

# DE L'OBSERVATION DE DONNÉES À LEUR TRAITEMENT PAR L'ANALYSE STATISTIQUE IMPLICATIVE : QUELQUES CONSTATS

Christian PELLOIS<sup>1</sup>

## TITLE

**From the observation of data to processing by implicative analysis: some established facts**

## RÉSUMÉ

Différentes formes de distribution de données témoignent des rapports variés qui peuvent exister entre des observations faites en continuum sur deux dimensions. Ces données peuvent être soumises au traitement statistique par le logiciel CHIC. Ce texte se propose d'examiner ces différentes formes de distribution de données. Il se propose, ensuite, de confronter l'interprétation qui peut être faite de ces formes, en termes de liaisons possibles entre les dimensions considérées, aux résultats obtenus par l'analyse statistique. Cette confrontation amène à envisager des règles, en termes de méthodologie de recherche, en particulier, lorsque CHIC est utilisé.

*Mots-clés : formes, distribution, données, continuum, analyse, statistique, implication.*

## ABSTRACT

Different forms of data distribution demonstrate the different relationships which can exist between observations made on a continuum in two dimensions. These data can be subjected to statistical processing by CHIC software. This text proposes to examine these different forms of data distribution. It then proposes to compare and contrast the interpretations which can be made of these forms in terms of the possible relationships between the dimensions resulting from statistical analysis. This leads to the consideration of rules for research methods, particularly when CHIC is used.

*Keywords: forms, distribution, data, continuum, analysis, statistics, implications.*

## 1 Introduction

Incontestablement l'analyse implicative, mise en œuvre par le logiciel CHIC, constitue une avancée importante dans le développement des traitements statistiques automatisés de données d'observation.

Ceci étant dit, une chose est d'élaborer des méthodes de traitement statistique sophistiquées tel le logiciel CHIC, autre chose est de les mettre en œuvre de manière rigoureuse et judicieuse. Ce texte a pour objectif de montrer que cela ne va pas de soi, et que cela pose des questions en termes d'interprétation, donc de sens qui peut être donné aux résultats obtenus, enfin, que cela nécessite de la part du chercheur, particulièrement

---

<sup>1</sup> 7 rue des pins, Courseulles sur mer, 14470 [Pellois.christian@wanadoo.fr](mailto:Pellois.christian@wanadoo.fr)

en sciences humaines, des qualités d'analyse, de sagacité et de rigueur, qu'il conviendrait de préciser.

Plus concrètement de quoi s'agit-il ? Le logiciel CHIC, de par sa conception de l'analyse statistique très élaborée peut amener le chercheur à penser qu'il maîtrisera aisément le sens des résultats fournis par cette méthode de traitement de données. Ce point de vue soulève cependant un certain nombre de questions dont les réponses pourraient peut-être permettre une utilisation certes plus prudente mais aussi plus pertinente, plus sûre de ce logiciel. C'est, modestement, par des exemples modélisés mais concrets, ce que ce texte souhaite éclairer. La difficulté tient plus particulièrement à un des points considéré comme extrêmement novateur et pertinent de ce traitement statistique par lequel, à côté d'un traitement en termes de similarité, il est proposé un traitement, en termes implicatif, d'une asymétrie que peut révéler la distribution des données.

Sur ce plan de l'interprétation des résultats fournis, la question du « *contexte* » dans lequel cette interprétation est faite paraît être une question fondamentale. C'est celle qui sera au centre des développements de la réflexion qui va suivre. Ce problème du *contexte* est souvent un problème difficile à aborder en sciences humaines, quand il n'est pas éludé purement et simplement.

En effet, l'analyse implicative s'appuie sur une conception de l'opération logique formelle qui lui correspond dont le *contexte* est la logique formelle et/ou les mathématiques traitant de cette logique formelle ou l'utilisant.

Mais ce *contexte* n'est déjà plus le même si la « *logique* » de l'implication est formulée dans un cadre particulier d'application. Par exemple, le fait d'appliquer cet *opérateur logique* de l'implication au fonctionnement d'un circuit électrique simple par lequel deux lampes A et B sont montées en série, peut amener à poser des règles de la façon suivante : « *la lampe A ne peut s'allumer que (ou s'allume si et seulement) si la lampe B s'allume* ». Dans ce cas de figure, raisonner logiquement renvoie à un sens « *concret* », « *situé* » donc « *local* » alors qu'au plan formel, la logique implicative se rapporte à une opération très générale relevant strictement d'un raisonnement abstrait. Le raisonnement appliqué réfère à certaines fonctionnalités du système électrique dont la compréhension relève des lois et théories de la physique appliquée. Elle est donc « *contextualisée* ». Enfin, la situation retenue est « *contextualisée* » de manière simple dans un système, conçu comme « *fermé* », et donc nécessairement isolé. Toute ouverture du système dans un système électrique plus large pourrait remettre en cause l'implication même, affirmée ici.

Mais traiter ainsi de l'implication sur des propositions concrètes c'est encore traiter de l'implication en tant qu'opération logique. L'analyse de données, au plan statistique traite, elle, plutôt, *des observations* faites, permettant de voir dans quelle mesure ces observations entrent dans une logique implicative. Passer de la logique de l'implication appliquée à des situations concrètes, à des observations qui peuvent en rendre compte, réfère à un *contexte* de nouveau différent.

Ensuite, un nouveau *contexte*, peut renvoyer aux difficultés méthodologiques à mettre en œuvre des observations. Ces difficultés peuvent amener à de l'incertitude, voire à des erreurs possibles sur la mesure, c'est-à-dire l'observation. Dans ce cas, la fiabilité tient à la répétition des mesures sur un même système, et/ou sur plusieurs

systèmes ceci, par exemple, pour tenter de montrer le caractère général, voire universel du principe soumis à la vérification sur le mode implicatif.

Du fait de l'incertitude de la mesure, les résultats obtenus en termes d'observation peuvent ne pas respecter dans leur totalité la logique implicative (par exemple, dans le cas des lampes, pour une raison quelconque, la lampe qui doit s'allumer est momentanément défectueuse et ne s'allume pas). Aussi il est accepté, notamment en sciences humaines, que le principe d'une liaison (ici implicative) est reconnue comme avérée si l'ensemble des observations concordent avec l'hypothèse de liaison émise, ceci à quelques exceptions près. Il s'agira alors de traiter d'une liaison *quasi implicative* et non plus de la logique implicative au sens strict. Cette règle du « *quasi-implicatif* » constitue un nouveau *contexte* à prendre en compte dans la démarche interprétative des résultats.

Enfin, les données recueillies, et là encore, le plus souvent, en sciences humaines, ne réfèrent pas à une logique de « *tout ou rien* » de type oui/non ou 0 ou 1 mais résultent d'observations faites sur un *continuum* renvoyant à des échelles de mesure, par exemple, de traits (de personnalité, de style cognitif, etc.), réfèrent à une logique « *plus ou moins oui* » ou « *plus ou moins non* ». Ceci constitue le cadre, encore, d'un nouveau *contexte* dans lequel se fait l'interprétation des résultats du traitement statistique.

Enfin, la population des observations considérées peut relever de sujets différents, auquel cas nous sommes, par ce nouveau *contexte*, au cœur des difficultés de l'interprétation des résultats fournis par l'analyse implicative.

Pour conclure cette introduction, toutes ces caractéristiques, prises globalement, contribuent à rendre plus complexes les situations traitées et à augmenter les risques en termes de pertinence d'interprétation des résultats du traitement statistiques.

Aussi ce texte a pour objet l'étude exploratoire de cette question de l'interprétation des résultats en fonction des *contextes*. Pour traiter de cette question, une approche particulière a été choisie. Cette approche peut être qualifiée de « *périphérique* » au regard de l'outil de traitement statistique ici considéré : le logiciel CHIC. En effet, il ne sera pas question de mathématiques au sens habituel et technique du terme, mais de l'observation de « *formes* » prises concrètement par des distributions de données, formes qui peuvent apparaître dans le domaine des sciences humaines. Les différentes formes de distribution de données seront présentées dans des tableaux modélisés croisant les observations faites sur deux dimensions quelconques (le plus souvent A et B). Le nombre de ces observations, faites sur chacune de ces deux dimensions, seront de l'ordre d'une centaine. Il sera ensuite question de ce que cette observation des formes de distribution amène comme réflexions concernant les interprétations qui peuvent en être données. Puis une analyse statistique de ces distributions d'observations sera effectuée à l'aide du logiciel CHIC. Les résultats obtenus sont, in fine, confrontés aux significations, plus ou moins hypothétiques, formulées préalablement au regard des formes de distribution présentées.

## 2 Différentes formes de présentation de l'analyse implicative et conceptions associées : bref rappel

### 2.1 L'implication logique prise au sens strict portant sur une population d'observations

Dans cette première présentation, très simple, il ne s'agit pas d'examiner seulement quatre propositions jugées comme fausses ou vraies et leurs rapports à une logique d'implication, mais *d'examiner, la manière dont des données recueillies par l'observation s'ordonnent au regard de cette logique là*. Nous passons, alors, comme cela a été évoqué ci-dessus, de la logique implicative, dûment formalisée, au champ de l'observation, de l'expérimentation. L'objectif n'est plus de rendre compte d'un raisonnement, ici le raisonnement implicatif, par une formalisation quelconque mais, comme le plus souvent en sciences humaines, de *l'adéquation entre ce raisonnement et la manière dont s'ordonne une population de données issues de l'observation* réalisée dans des contextes plus ou moins spécifiés.

Dans le cas de figure de la multiplication des observations concernant l'implication de type si A alors B, formulée, par exemple, à propos d'un circuit électrique composé de lampes, installées en série, susceptibles de s'allumer ou de s'éteindre, les résultats obtenus, compatibles avec le modèle de l'implication, portant sur 99 observations, peuvent se répartir de la manière suivante (Cf. Tableau 1 ci-dessous).

TABLEAU 1 – Distribution de données compatible avec le modèle de l'implication

A allumée (P)	0	33
A éteinte (non P)	33	33
	B éteinte (non Q)	B allumée (Q)

Cette présentation de résultats d'observation correspond rigoureusement à ce qui peut être défini en termes d'implication stricte.

Mais l'analyse implicative, et le logiciel CHIC qui lui est associé, porte non pas sur l'implication stricte mais sur la « *quasi-implication* ». Dans le cas figure de la quasi-implication il est toléré qu'apparaissent quelques observations dans la case, ici à effectif nul. Ces observations seront renvoyées, le plus souvent, à l'incertitude de la mesure et aux erreurs d'observation, toujours possibles, notamment, en sciences humaines.

### 2.2 La « *quasi-implication* » portant sur un continuum de scores variés

#### 2.2.1 Un continuum de scores et ce qui en résulte du point de vue de la question de l'implication

Le traitement par l'analyse implicative, comme présenté au point 2.1, peut porter sur des données de type dichotomique, à savoir oui/ non, *présence* ou *absence* d'un état, 0 ou 1. Cette répartition des données entre quatre positions clairement définies n'est qu'un

cas particulier de répartition, notamment dans le domaine des sciences humaines. Dans ce domaine et en particulier celui de la psychologie, le plus souvent, les données se distribuent sur un continuum de performances, ou de scores, traitant de caractéristiques opposées (par exemple : extraversion/introversion, *attribution* interne/externe, etc.) ou traitant de la présence plus ou moins accentuée d'une caractéristique ou d'une autre (l'émotivité, une capacité de mémoire, etc.). Les scores obtenus par une population considérée présentent une certaine dispersion et se répartissent alors sur une échelle allant d'une valeur (*x*) à une valeur (*y*). Dans le cadre de l'utilisation du logiciel CHIC ces valeurs sont ramenées à une échelle standard de scores allant de 0 à 1(0.10, 0.23, 0.67, etc.) ceci afin de permettre le traitement conjoint de l'ensemble des scores obtenus sur les différentes dimensions retenues pour un traitement d'ensemble de ces données par le logiciel CHIC.

Sur la base de ce continuum, la distribution susceptible d'être compatible avec la logique implicative, des scores obtenus sur deux dimensions, peut se faire là aussi, lorsque l'on croise les données, selon un ordonnancement globalement analogue à celui du tableau 1 (Cf. Tableau 2 ci-dessous).

TABLEAU 2 – Distribution de données « proches » de la précédente distribution mais établi sur un continuum de scores

	5	5	5	5	5	15	15	15	15	15	100
0,9	0	0	0	0	0	1	1	1	1	1	5
0,8	0	0	0	0	0	1	1	1	1	1	5
0,7	0	0	0	0	0	1	1	1	1	1	5
0,6	0	0	0	0	0	1	1	1	1	1	5
0,5	0	0	0	0	0	1	1	1	1	1	5
0,4	1	1	1	1	1	2	2	2	2	2	15
0,3	1	1	1	1	1	2	2	2	2	2	15
0,2	1	1	1	1	1	2	2	2	2	2	15
0,1	1	1	1	1	1	2	2	2	2	2	15
0	1	1	1	1	1	2	2	2	2	2	15
A ↑	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	100
B →											

Que dire de ce tableau ? A l'évidence<sup>2</sup>, si le tableau présente une distribution des données dont l'allure générale semble comparable au tableau 1, il ne peut lui être assimilé, et cela au moins pour deux raisons.

a) Par l'acceptation d'observations qui s'inscrivent sous la forme d'un continuum, nous passons d'une logique de l'implication à une logique « *d'implication relative* » ou « *graduée* » ou à une logique de « *tendance implicative* ».

<sup>2</sup> Une autre évidence peut être formulée. Par rapport à un schéma dichotomique simple (oui/non, 0 ou 1) le fait d'introduire un continuum sur une échelle de scores variés et progressifs fait passer d'un système présentant un degré de liberté limité à un système présentant un degré de liberté plus grand du fait du cadre même retenu pour traiter des informations recueillies à savoir un tableau à quatre cases auquel se substitue, dans le cas de scores répartis sur une échelle à dix graduations, un tableau de cent cases.

b) sur le plan de la distribution des données, et au regard de ce principe « *modifié* » de l'implication stricte (valeurs de B supérieures à valeurs de A), une part non négligeable des observations présentées dans le tableau ci-dessus, ne respecte pas ce principe. Elles correspondent plutôt au *schéma implicatif inverse* (valeurs de A supérieures à valeurs de B).

Pour qu'il y ait l'observation d'une implication systématique, mais variable dans son caractère radical, il conviendrait que l'ensemble des données se répartissent, par exemple, tel que présenté dans la schématisation ci-dessous (Cf. Tableau 3).

TABLEAU 3 – Distribution sur un continuum de scores globalement compatible avec la logique de l'implication

	<b>0</b>	<b>3</b>	<b>6</b>	<b>8</b>	<b>10</b>	<b>12</b>	<b>14</b>	<b>14</b>	<b>16</b>	<b>17</b>	<b>100</b>
0,9	0	0	0	0	0	0	0	0	0	0	0
0,8	0	0	0	0	0	0	0	0	0	3	3
0,7	0	0	0	0	0	0	0	0	3	3	6
0,6	0	0	0	0	0	0	0	3	3	2	8
0,5	0	0	0	0	0	3	3	2	2	2	10
0,4	0	0	0	0	3	3	2	2	2	2	12
0,3	0	0	0	3	3	2	2	2	2	2	14
0,2	0	0	3	3	2	2	2	2	1	1	15
0,1	0	3	3	2	2	2	1	1	1	1	15
0	3	3	2	2	2	2	1	1	1	1	17
<b>A</b> ↑	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	<b>100</b>
<b>B</b>	→										

Notons sur le tableau ci-dessus une caractéristique portant sur l'allure générale des deux distributions. Ces deux distributions sont nettement asymétriques, mais de façon différente. En ce qui concerne A, le regroupement de la majorité des observations se fait vers des scores faibles. Pour B le regroupement de la majorité des observations se fait vers des scores élevés. Cette particularité sera traitée plus complètement par la suite.

### 2.2.2 La question de la « quasi-implication »

Toujours dans le domaine des sciences humaines, les données se présentent, le plus souvent, dans une certaine dispersion par rapport au *principe même d'ordonnement des données retenues*. Cette dispersion traduit, alors, non pas l'implication stricte entre deux dimensions mais une *quasi-implication* (Gras, 2005, p. 12). Cette position divergente au regard du modèle de l'implication stricte est cependant reconnue comme acceptable pour traiter de systèmes complexes, notamment du fait, comme cela a déjà été évoqué, de la part d'incertitude sur la mesure, particulièrement en sciences humaines<sup>3</sup>.

<sup>3</sup> Fidélité plus ou moins élevée de la mesure, donc part d'erreur de mesure que peut receler une mesure ponctuelle, fluctuations ponctuelle des résultats tenant à des aspects d'une certaine instabilité humaine et de ses contextes, de l'instabilité, des fluctuations des équilibres dynamiques, qui en découle, pour les

A partir d'une conception référant plus à une *tendance implicative* observée sur un continuum, et l'acceptation de « *dérogations* » tolérées par ce nouveau modèle formel, la distribution des observations pourrait, par exemple, être le suivant (Cf. Tableau 4).

TABLEAU 4 – Distribution de données présentant une situation de quasi-implication

	0	3	7	11	14	12	13	14	14	12	100
0,9	0	0	0	0	0	0	0	0	0	0	0
0,8	0	0	0	0	0	0	0	0	0	3	3
0,7	0	0	0	0	0	0	0	1	3	3	7
0,6	0	0	0	1	1	0	1	3	3	2	11
0,5	0	0	0	1	1	2	3	3	2	2	14
0,4	0	0	0	0	2	3	3	2	2	2	14
0,3	0	0	0	1	3	3	2	2	2	0	13
0,2	0	0	1	3	3	2	2	2	2	0	15
0,1	0	0	3	3	2	2	2	1	0	0	13
0	0	3	3	2	2	0	0	0	0	0	10
A	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	100
B											

### 3 Différentes formes caractéristiques modélisées de données susceptibles d'apparaître

Les distributions de données telles que cadrées ci-dessus peuvent, cependant, donner lieu à de multiples autres formes de distributions à commencer par celles les plus classiquement observées en sciences humaines, jusqu'à des formes plus complexes. C'est ce qui va être examiné maintenant.

#### 3.1 Des formes classiques de distribution de données

Les principes de continuum de distribution de scores et la « *quasi implication* » permettent qu'apparaissent, bien sûr, des répartitions de données traduisant une certaine *tendance implicative* comme ce qui est présentés sur le tableau 4 ci-dessus, mais aussi des répartitions plus classiques en termes de *corrélacion*, ou de *corrélacion négative* traduisant une possible « *incompatibilité* » entre deux dimensions. Il est à noter sur ce point que le passage à l'incompatibilité « *vrai* » mériterait bien d'autres développements, comme en ce qui concerne l'implication « *vrai* »<sup>4</sup>. En effet, l'incompatibilité prend plus pleinement son sens logique si à des scores croissants pour une variable A correspond des scores décroissants d'une variable B et que l'on observe également qu'à l'accroissement des scores de B correspond une décroissance des scores

dimensions examinées, etc.

<sup>4</sup> Sur la question de l'implication, voir, par exemple, l'importance qu'il y aurait à prendre en compte la contraposée (Gras, Couturier, 2010, p. 191).

de A. Ces formes de distribution classiques caractérisant la corrélation positive ou une forme d'incompatibilité, entre deux dimensions peuvent se présenter de la manière suivante (Cf. Tableau 5 et 6)<sup>5</sup>.

TABLEAU 5 – Distribution de données sur le modèle de la co-variation entre deux dimensions

	1	5	9	15	20	21	14	9	5	1	100
0,9	0	0	0	0	0	0	0	0	0	1	1
0,8	0	0	0	0	0	0	1	1	3	0	5
0,7	0	0	0	0	0	2	2	4	1	0	9
0,6	0	0	0	0	2	3	6	2	1	0	14
0,5	0	0	0	3	4	9	3	2	0	0	21
0,4	0	0	2	3	9	4	2	0	0	0	20
0,3	0	1	2	6	3	3	0	0	0	0	15
0,2	0	1	4	2	2	0	0	0	0	0	9
0,1	0	3	1	1	0	0	0	0	0	0	5
0	1	0	0	0	0	0	0	0	0	0	1
A ↑	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	100
B →											

TABLEAU 6 – Distribution de données sur le modèle d'une possible « incompatibilité » entre deux dimensions

	1	5	9	15	20	21	14	9	5	1	100
0,9	1	0	0	0	0	0	0	0	0	0	1
0,8	0	3	1	1	0	0	0	0	0	0	5
0,7	0	1	4	2	2	0	0	0	0	0	9
0,6	0	1	2	6	3	3	0	0	0	0	15
0,5	0	0	2	3	9	4	2	0	0	0	20
0,4	0	0	0	3	4	9	3	2	0	0	21
0,3	0	0	0	0	2	3	6	2	1	0	14
0,2	0	0	0	0	0	2	2	4	1	0	9
0,1	0	0	0	0	0	0	1	1	3	0	5
0	0	0	0	0	0	0	0	0	0	1	1
A ↑	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	100
B →											

<sup>5</sup> NB : le traitement par CHIC, au moins dans le premier cas de figure, relève de la recherche de similarité entre les dimensions considérées. Par contre, la recherche d'une relation en termes d'implication ne semble pas devoir se poser ici.

Mais il peut apparaître bien d'autres formes théoriques de distributions conjointes de données. Certaines pouvant renvoyer plus ou moins explicitement, ou pas, à des formes plus ou moins proches de celles des tendances implicatives telles que présentées dans les tableaux 2, 3, 4.

### 3.2 La forme de distribution de données peut renvoyer à une interprétation selon laquelle une série de mesures porte sur des exercices plus « difficiles » que ceux de l'autre série

Dans l'observation empirique directe de données, il peut apparaître que les fréquences des scores obtenus se traduisent par des distributions différentes, plus ou moins asymétriques, d'une dimension à une autre. La question est de savoir ce que peut sous-tendre, en termes de sens, de telles formes de distributions, puis de voir, ensuite, comment elles sont susceptibles d'être prises en compte dans CHIC. Deux exemples simples seront abordés ici.

Le premier cas de figure considéré est celui d'une dimension A dont les scores ont leurs fréquences les plus élevées proches de la valeur minimale (ramenée à 0 sur l'échelle standard) et d'une dimension B dont les fréquences se répartissent de manière monotone sur l'échelle allant de 0 à 1.

Ceci nous amène à une distribution de données qui peut se présenter ainsi (Cf. Tableau 7 ci-dessous).

TABLEAU 7 – Distribution de données traitant d'une épreuve « facile » et d'une épreuve « adaptée » à la population considérée

	<b>21</b>	<b>20</b>	<b>20</b>	<b>18</b>	<b>15</b>	<b>2</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>100</b>
0,9	1	1	1	1	1	1	1	1	1	1	<b>10</b>
0,8	2	2	2	2	1	1	0	0	0	0	<b>10</b>
0,7	2	2	2	2	2	0	0	0	0	0	<b>10</b>
0,6	2	2	2	2	2	0	0	0	0	0	<b>10</b>
0,5	2	2	2	2	2	0	0	0	0	0	<b>10</b>
0,4	2	2	2	2	2	0	0	0	0	0	<b>10</b>
0,3	2	2	2	2	2	0	0	0	0	0	<b>10</b>
0,2	2	2	2	2	2	0	0	0	0	0	<b>10</b>
0,1	3	2	2	2	1	0	0	0	0	0	<b>10</b>
0	3	3	3	1	0	0	0	0	0	0	<b>10</b>
<b>A</b> ↑	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	<b>100</b>
<b>B</b> →											

L'interprétation, en termes de sens, concernant ces deux distributions croisées, par exemple s'agissant de performances cognitives ou d'exercices de grammaire ou de mathématiques, peut être la suivante. A savoir : pour une population donnée et au regard des fréquences différentielles, par tranches de scores, sur B, la série d'exercices portant sur la dimension B peut être considérée comme « plus difficile » que la série d'exercices portant sur la dimension A.

### 3.3 Une autre forme de distribution des données peut renvoyer à une interprétation en termes d'exercices plus « difficiles » pour l'une des séries (B) et plus « faciles » pour l'autre (A)

Le cas de figure traité ici est celui de la même dimension A que ci-dessus et d'une dimension B dont les scores, au contraire de ceux de la dimension A, ont leurs fréquences les plus élevées proches des valeurs maximales. Ce qui nous amène à une distribution de données qui se présente tel que ci-dessous (Cf. Tableau 8).

TABLEAU 8 – Distribution de données traitant de deux épreuves « faciles » concernant la population considérée

	21	20	20	18	15	2	1	1	1	1	100
0,9	4	4	4	2	4	1	1	1	0	0	21
0,8	4	4	4	4	4	0	0	0	0	0	20
0,7	4	4	4	4	4	0	0	0	0	0	20
0,6	4	4	4	4	2	0	0	0	0	0	18
0,5	2	4	4	4	1	0	0	0	0	0	15
0,4	2	0	0	0	0	0	0	0	0	0	2
0,3	1	0	0	0	0	0	0	0	0	0	1
0,2	0	0	0	0	0	0	0	0	1	0	1
0,1	0	0	0	0	0	1	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0	1	1
A ↑	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	100
B →											

Cette nouvelle distribution, du fait de la présence de la tendance « facile » pour la seconde dimension, la première traduisant toujours une dissymétrie de distribution dans le sens « difficile », amène nécessairement une accentuation de la fréquence des observations sur toute une partie du tableau. Cette accentuation se traduit par une apparence, référant à une possible « tendance implicite » alors que l'interprétation en terme de « plus facile/ plus difficile » pourrait être la seule valide.

### 3.4 Une forme susceptible de traduire des effets de seuil ?

Une forme différente mais potentiellement assimilable aux cas précédents peut caractériser des effets de « seuil ». De quoi s'agit-il ? Dans le domaine des sciences que ce soit en sciences physiques (diodes) ou en sciences de la vie (cycle ovulaire), comme dans celui des sciences humaines, l'émergence d'une qualité (une sécrétion hormonale, l'apparition d'un fait, etc.) peut n'apparaître qu'à un certain niveau d'expression d'une autre qualité. Ce niveau de déclenchement peut être appelé « seuil » d'émergence. En psychologie, cela peut être le cas, par exemple, d'une capacité « d'espace mental », concept référant à J. Pascual Leone (Ribeupierre, 1980), qui, à un certain niveau de capacité permettrait le développement opératoire logique de l'enfant, au sens piagétien du terme. Si le phénomène peut donner lieu à un traitement quantitatif, l'expression quantitative de l'effet « seuil », peut apparaître sous la forme suivante (Cf. Tableau 9).

TABLEAU 9 – Distribution de données traitant de la notion de « seuil »

	0	0	0	1	1	8	14	24	26	26	100
0,9	0	0	0	0	0	0	0	0	0	9	9
0,8	0	0	0	0	0	0	0	6	7	6	19
0,7	0	0	0	0	0	1	2	4	6	5	18
0,6	0	0	0	0	0	2	3	3	4	5	17
0,5	0	0	0	0	0	1	2	3	4	1	11
0,4	0	0	0	0	0	1	2	3	3	0	9
0,3	0	0	0	0	0	1	2	3	2	0	8
0,2	0	0	0	0	0	1	2	2	0	0	5
0,1	0	0	0	0	0	1	1	0	0	0	2
0	0	0	0	1	1	0	0	0	0	0	2
A	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	100
B											

La forme de la distribution des données peut faire penser, comme précédemment, à l'expression d'une « *tendance implicative* ». Mais, à la différence du cas de figure précédent, il existe bien derrière l'effet « seuil », une logique *implicative*. En effet, l'expression de scores sur une dimension, ici A, dépend bien de l'évolution des scores sur une dimension B, et ceci à partir d'un certain score de B. En ce sens, les données du tableau vont bien dans le sens attendu puisqu'il existe un maximum d'observations du côté de l'axe diagonal référent à la logique implicative retenue (65 observations) contre un minimum dans le sens inverse (11 observations), 24 observations se plaçant, elles, sur la diagonale.

### 3.5 Une forme traduisant de possibles effets en termes d'« optimalité »

Le principe de l'*optimalité* consiste à dire qu'une variable s'exprime, non pas pour des valeurs extrêmes d'une autre variable (très faibles ou très élevées) mais pour des valeurs intermédiaires, par exemple moyennes, de cette variable. Cette situation d'optimalité est loin d'être rare en sciences humaines, en psychologie en particulier. Cela peut concerner par exemple l'émotion qui, à un niveau d'intensité modérée, peut améliorer des capacités cognitives (attention, vigilance, mobilisation sur une tâche, mémoire, etc.) mais, qui, à un niveau d'intensité élevée, a des effets plus perturbants que favorables. Il en va de même des conduites éducatives des parents au regard, par exemple du développement opératoire (Lautrey, 1980) des enfants, en particulier, et plus généralement au regard de l'efficacité adaptative. Les recherches sur ces questions suggèrent que les conduites éducatives trop « *autoritaires* » (potentiellement « *rigides* ») ou trop « *laisser faire* » (potentiellement « *laxistes* ») auraient des effets négatifs alors que des conduites éducatives relevant davantage de modalités « *mixtes* » et donc plus *souples* et *variées*, en fonction des situations rencontrées, aurait des effets plus positifs.

La situation d'optimalité peut être présentée sous différentes formes. Il sera proposé, ici, une seule de ces formes (Cf. Tableau 10 ci-dessous).

TABLEAU 10 – Distribution de données traitant de la notion « d'optimalité »

	6	12	17	7	8	8	7	17	12	6	100
0,9	0	0	0	7	8	8	7	0	0	0	30
0,8	0	0	7	0	0	0	0	7	0	0	14
0,7	0	0	6	0	0	0	0	6	0	0	12
0,6	0	0	4	0	0	0	0	4	0	0	8
0,5	0	5	0	0	0	0	0	0	5	0	10
0,4	0	4	0	0	0	0	0	0	4	0	8
0,3	0	3	0	0	0	0	0	0	3	0	6
0,2	3	0	0	0	0	0	0	0	0	3	6
0,1	2	0	0	0	0	0	0	0	0	2	4
0	1	0	0	0	0	0	0	0	0	1	2
A	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	100
B											

Sur le plan de la logique, il existe bien, là aussi, une liaison qui pourrait s'interpréter en termes d'implication logique entre la variable A et la variable B, de type B « est nécessaire » à A (A « implique » B ?) dans la mesure où l'augmentation de B est associée à une évolution de A dans un sens croissant puis dans un sens décroissant. Pour autant l'inverse, quoique moins évident, pourrait aussi s'interpréter en ce sens (des fluctuations de A entraînent une évolution croissante et continue des scores de B... ?). Mais la situation est ici plus complexe puisque l'évolution de A en fonction de B n'est pas linéaire. Le traitement par CHIC peut, cependant, donner un résultat en termes d'indice implicatif. Cet indice peut caractériser une implication dans le sens le moins évident (B implique A). Cela pourrait tenir au fait que l'effectif de la population le plus important se trouve dans la partie supérieure gauche par rapport à la diagonale du rectangle<sup>6</sup>. Ceci étant dit, s'il y a moins d'observations B dans une logique d'implication inverse cela représente tout de même près du quart de la population totale... En fait s'il existe bien un type d'implication, il ferait référence, à la différence de la logique trop simpliste ci-dessus, à un ordonnancement des données qui ne relèverait pas de leur répartition telle que ci-dessus et de la forme de courbe d'« optimalité » qui lui correspondrait. A propos du rôle des conduites éducatives, par exemple, la représentation correcte relevant de l'implication réelle serait la suivante (Cf. Tableau 11, ci-dessous).

<sup>6</sup> 71 contre 22 dans la partie inférieure droite, 7 étant dans la zone médiane.

TABLEAU 11 – Nouvelle distribution de données plus compatible avec la logique implicative, traitant de la notion « d'optimalité »<sup>7</sup>

Niveau de développement opératoire <u>élevé</u>	0	z
<u>Faible</u> niveau du développement opératoire	x	y
	Conduites éducatives extrêmes (« autoritaire » / "laisser faire")	Conduites éducatives mixtes et variées, en fonction des situations rencontrées

Les valeurs dichotomiques sont, ici, en ce qui concerne le développement opératoire, un indice de développement faible et un indice de développement élevé<sup>8</sup>. En ce qui concerne les pratiques éducatives, une position considérée, par hypothèse, comme peu pertinente, voire incompatible, avec le développement opératoire de l'enfant regrouperait les positions extrêmes des scores relevés<sup>9</sup>. L'autre position retiendrait les scores moyens référant à une conduite davantage « mixte » et plus alternative en fonction de situations rencontrées (c'est-à-dire, ici, les scores moyens). Dans ce cas de figure les fréquences des observations recueillies ( $x$ ,  $y$ ,  $z$ ) en accord strict avec la logique implicative ci-dessus présentée se répartiraient tel que dans le tableau 11.

Une remarque : cette présentation montre bien la nécessité, pour le chercheur, de développer des qualités d'analyse préalable, une perspicacité certaine, aussi bien dans la réflexion initiale et la préparation de sa recherche que dans sa mise en œuvre, qualités bien utiles également dans une utilisation circonstanciée, réfléchie, du logiciel CHIC.

### 3.6 Une forme « complexe » : vers des approches différentielles ?

Dans les formes précédentes, à quelques observations près, les situations représentées relèvent de tendances simples parce qu'univoques. Les observations qui contredisent ou même simplement atténuent la tendance dans la forme dominante de répartition des données sont, le plus souvent, très largement minoritaires voire marginales. Qu'en est-il de situations plus complexes caractérisées par le fait que plusieurs tendances sont susceptibles d'apparaître dans une distribution de données et ceci dans des proportions plus ou moins équilibrées. Ce problème donne lieu, ici, à un traitement en plusieurs étapes. Une première option possible consiste à observer une *tendance implicative* pour une partie des valeurs d'une variable (B), par exemple, les valeurs supérieures, et à observer une *tendance implicative* inverse pour la partie

<sup>7</sup> Mais les cas d'observation  $Y$ , naturellement, n'apparaissent pas ici dans la distribution des données correspondant à la notion d'optimalité illustrée par le tableau 10. Le tableau plus complet en ce sens relève d'une distribution croisée comparable mais plus complexe.

<sup>8</sup> Par exemple : la position au dessus (développement élevé) ou au dessous (développement faible) d'un certain seuil caractéristique de scores de performances cognitives de l'enfant sur cette dimension du développement cognitif.

<sup>9</sup> Les positions style éducatif « autoritaire » et style éducatif « laisser faire ».

inférieure des valeurs de cette même variable (B). Cela donne le tableau suivant (Cf. Tableau 12 ci-dessous).

TABLEAU 12 – Distribution complexe de données traitant de l'inversion de la tendance implicative en fonction des scores sur une des deux dimensions

	4	7	9	12	15	17	14	11	7	4	100
0,9	0	0	0	0	0	3	3	3	3	4	16
0,8	0	0	0	0	0	4	4	4	4	0	16
0,7	0	0	0	0	0	3	3	4	0	0	10
0,6	0	0	0	0	0	3	4	0	0	0	7
0,5	0	0	0	0	0	4	0	0	0	0	4
0,4	0	0	0	0	4	0	0	0	0	0	4
0,3	0	0	0	4	3	0	0	0	0	0	7
0,2	0	0	4	3	3	0	0	0	0	0	10
0,1	0	4	2	2	2	0	0	0	0	0	10
0	4	3	3	3	3	0	0	0	0	0	16
A ↑	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	100
	B →										

Le tableau ainsi construit montre une répartition presque équilibrée entre une tendance à l'implication dans un sens et une tendance à l'implication dans l'autre sens. Nous pourrions d'ailleurs avoir une position parfaitement équilibrée entre les deux sens si le choix n'avait pas été fait volontairement de déséquilibrer dans un sens la répartition des effectifs au détriment de l'autre sens, ceci par le choix volontaire de favoriser légèrement une des deux tendances. Sur un plan formel il existe donc bien deux tendances implicatives presque équivalentes, en jeu dans les données présentées. Une tendance implicative de type A implique B pour des valeurs de A égales ou inférieures à 0,4 et une tendance implicative de type B implique A pour des valeurs de A égales ou supérieures à 0,6.

Une telle observation a pu être faite dans le cadre d'une recherche récente (Chamsine, 2012, à paraître). Quel sens donner à une telle répartition ? Un sens possible est celui d'un **changement** qui s'opèrerait<sup>10</sup> plus ou moins brusquement, sur une dimension, en fonction de l'évolution des scores sur l'autre dimension<sup>11</sup>, changement se traduisant par une inversion de la tendance implicative. L'hypothèse formulée ici est que dans un tel cas de figure, le chercheur, s'il s'en tient aux indications fournies par une analyse implicative globale, sera amené à conclure que pour les deux variables considérées, la *tendance implicative* fonctionne (peut-être) dans un sens, du fait de la légère accentuation dans ce sens du poids quantitatif des observations, alors même que cette affirmation est manifestement erronée concernant les résultats obtenus pour une partie des données.

<sup>10</sup> Dans le comportement, dans les attitudes, dans les stratégies de fonctionnement des sujets, ou dans les mécanismes d'action au regard des variables considérées pour elles-mêmes, etc.

<sup>11</sup> Performances cognitives, niveau de comportement, plus ou moins grande accentuation d'un trait de personnalité, etc.

Une suggestion méthodologique peut alors être brièvement présentée ici, et reprise dans une approche plus globale ultérieurement. Dans ce cas de figure, le chercheur peut avoir quelques indices. Il peut constater que les indices implicatifs fournis par CHIC sont proches. Si ces indices sont proches, il peut avoir recours à l'observation directe concrète de la distribution de ses données telle qu'elle est proposée dans ce texte (Cf. les tableaux ici présentés). A partir de là, il peut se rendre compte de la situation différentielle en fonction des scores sur une des deux dimensions. Il peut, alors, définir en fonction de ses observations, deux sous populations, et revoir dans l'analyse par CHIC, si le calcul rend compte de ces tendances implicatives différentes pour les deux sous groupes d'observations constitués. En fonction des valeurs des indices obtenus le chercheur prend alors la décision de reconnaître, ou pas, le résultat différentiel.

Notons, par ailleurs, que le tableau ci-dessus présente aussi une forme corrélationnelle en superposition de la double implication dépendant du niveau de score atteint sur la dimension A. Ce tableau pourrait prendre une forme non corrélationnelle et donc simplement « *doublement* » implicative.

### 3.7 Corrélation « *négative* », approche différentielle et implication : une forme susceptible de traduire des cheminements vicariants

Il convient d'en venir maintenant au fait que nous examinons, en sciences humaines, derrière les observations, des *sujets*<sup>12</sup> et que les caractéristiques humaines des sujets ouvrent à des complexités, susceptibles de poser de délicats problèmes d'analyse et d'interprétation des résultats. Pour aborder cela, d'autres formes de croisement de données peuvent être évoquées. Une de ces formes peut prendre l'allure générale d'une corrélation négative entre les dimensions considérées (Cf. Tableau 13 ci-dessous).

TABLEAU 13 – *Distribution complexe de données traitant de sous-groupes aux « profils » différents ou aux cheminements alternatifs différents (vicariance)*

	2	14	18	14	2	2	14	18	14	2	100
0,9	2	0	0	0	0	0	0	0	0	0	2
0,8	0	4	6	4	0	0	0	0	0	0	14
0,7	0	6	6	6	0	0	0	0	0	0	18
0,6	0	4	6	4	0	0	0	0	0	0	14
0,5	0	0	0	0	2	0	0	0	0	0	2
0,4	0	0	0	0	0	2	0	0	0	0	2
0,3	0	0	0	0	0	0	4	6	4	0	14
0,2	0	0	0	0	0	0	6	6	6	0	18
0,1	0	0	0	0	0	0	4	6	4	0	14
0	0	0	0	0	0	0	0	0	0	2	2
A ↑	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	100
B →											

<sup>12</sup> Notons que les sujets sont aussi bien des « *observateurs* » que des « *observés* ».

S'agit-il, simplement, ici, comme au point 3.1, d'une forme d'incompatibilité possible entre les deux dimensions ? Dans cette nouvelle distribution de données, d'autres hypothèses interprétatives pertinentes sont possibles. Considérons qu'il s'agit, là, d'une recherche en sciences humaines portant sur deux dimensions et concernant 100 sujets différents. Outre la tendance globale à une forme évoquant une corrélation négative, il peut être, aussi, fait l'hypothèse que la population considérée est constituée de deux sous-groupes relativement distincts. Ces deux groupes pourraient être caractérisés de deux manières différentes. L'un des groupes correspondrait à un profil « *type* » fait de scores<sup>13</sup> élevés sur une des dimensions considérées (A) et de scores faibles sur l'autre (B). L'autre groupe, à l'inverse, caractériserait un profil « *type* » fait de scores plutôt faibles sur la dimension A et de scores plus élevés sur la dimension B. Tout se passe comme si nous avions là une répartition effective de deux populations, aux *profils* contrastés (A-/B+ et A+/B-) traduisant des styles d'expression différents. Mais ces profils varient quelque peu, à l'intérieur de cette règle globale de dispersion différentielle des données, donnant alors, sur un continuum, la forme globale, plus ou moins caractéristique, d'une corrélation négative. En fait, à l'intérieur de cette forme globale, il s'agirait de deux distributions relativement indépendantes. Il n'y aurait donc pas, nécessairement ici de lien général à retenir, malgré la suggestion de la forme dans son ensemble. Nous aurions deux configurations, deux styles d'expression de groupes de sujets différents. Ces styles d'expression peuvent tenir à de multiples raisons : constitutives, acquises, relevant d'une position plus ou moins mixte, « *dispositionnelles* » (Bruchon-Schweitzer, 2002) tôt installées dans le développement psychologique des personnes, et pouvant renvoyer, par exemple, à des contextes culturels contrastés, ou même à des choix délibérés, aux intérêts de la personne, investie davantage dans une dimension plutôt que dans l'autre et vice versa pour l'autre sous groupe de personnes. Ces profils ou configurations sont fréquentes en psychologie cognitive (exemple des styles cognitifs). Ils peuvent d'ailleurs interférer avec le développement cognitif de la personne. C'est le cas du développement opératoire, au sens piagétien du terme, pris sous sa forme différentielle vicariante (Longeot, 1978). Le sujet peut avoir des performances plutôt spatiales/représentatives ou plutôt logico/verbales. Ces performances différentielles permettraient deux voies alternatives (vicariance) du développement opératoire de l'intelligence. En retour, une fois un niveau de développement atteint, celui-ci pourrait agir sur les deux dimensions qui lui ont permis de se développer. Dans ce cas de figure, il n'existerait ni, réellement, corrélation négative, pas plus qu'une quelconque éventuelle *incompatibilité* entre les deux dimensions, mais l'expression de deux styles de développement différents, aboutissant à terme, au même niveau du développement opératoire. Quant à la question de l'implication, qu'en est-il ? Dans ce cas de figure, si l'implication existe, elle existe pour chacune des deux sous-populations mais ne porte pas nécessairement sur chacune des deux dimensions considérées. Elle porte sur chacune des dimensions considérées via une troisième dimension : le développement opératoire. C'est donc le chercheur, qui, au regard du modèle théorique ici esquissé, dans *une recherche confirmatoire*, va décider des vérifications ciblées choisies qu'il va faire *au moyen du* traitement statistique des données lui-même choisi en fonction de ce qu'il cherche à vérifier (implication, similarité, etc.). Ceci n'exclut nullement *une recherche de type exploratoire*, mais il conviendrait alors de s'aider, dans ce cas, a minima, de

<sup>13</sup> Par exemple : sur une performance cognitive, un trait de personnalité, etc.

l'observation circonstanciée du tableau de distribution de données, observation faite en tenant compte du cadre conceptuel de la recherche considérée.

### 3.8 Des données recueillies dans une population à des sous-groupes de cette population, voire à chaque sujet observé

Il a été évoqué, dans un tableau de quasi-implication, présenté précédemment (Cf. section 2.2.2), le fait que des observations en petits nombre peuvent caractériser une tendance implicative inverse de celle correspondant à l'ensemble des observations. Néanmoins, la conclusion serait que l'hypothèse implicative directe peut être retenue, en considérant que les observations infirmant l'hypothèse peuvent relever, notamment, d'erreurs d'observation. Cependant, d'autres explications sont susceptibles d'être évoquées. Elles font référence, par exemple, comme ci-dessus, à des profils différents et à de possibles cheminements vicariants (Pellois, 2007, p. 64 et suivantes) ainsi qu'à des différences dans des cheminements observés concernant le développement de l'enfant ou des stratégies d'expression<sup>14</sup>. Peut-on aller plus loin dans ce type de raisonnement ? Pour aborder cela, il convient de regarder attentivement une forme de répartition des données tel que présenté sous la forme complexe suivante (Cf. Tableau 14 ci-dessous).

TABLEAU 14 – *Distribution de données renvoyant à une approche différentielle complexe*

	5	9	11	15	14	14	11	10	7	4	100
0,9	2	1	0	0	0	0	0	0	1	1	5
0,8	2	2	2	2	2	2	0	2	3	1	18
0,7	0	2	2	2	1	2	3	3	2	0	17
0,6	0	0	1	2	2	3	3	3	1	0	15
0,5	0	0	1	2	2	4	3	1	0	0	13
0,4	0	0	0	2	4	2	1	0	0	1	10
0,3	0	0	1	4	2	1	1	0	0	0	9
0,2	0	1	3	1	1	0	0	1	0	0	7
0,1	0	3	1	0	0	0	0	0	0	0	4
0	1	0	0	0	0	0	0	0	0	1	2
<b>A</b> ↑	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	100
<b>B</b> →											

Il est clair que cette forme peut présenter la caractéristique d'une certaine implication de type B implique A. Mais si cette forme se caractérise bien par une forme dominante elle se caractérise aussi par une certaine divergence par rapport à cette dominante, plus ou moins importante et sous des formes, elles mêmes variées différentes les unes par rapport aux autres. Doit-on, comme cela est classiquement le cas en sciences humaines, renvoyer les observations divergentes au regard d'une tendance « dominante » aux erreurs de mesure, à une part de fluctuation, d'incertitude ? Ou bien doit-on renvoyer à des styles ou des profils différents, par exemple, sources

<sup>14</sup> Cf. Pellois, 2010, p. 209 et, ici, la forme « complexe » évoquée à la section 3.6.

potentielles d'une certaine vicariance dans le développement, comme évoqué ci-dessus ?

L'observation attentive mais différentielle de cette distribution amène à considérer, tour à tour, la corrélation pour un groupe de sujets, une corrélation négative (*incompatibilité* ?) pour un autre groupe, une implication (Cf. ci-dessus) de type  $B \Rightarrow A$  pour une majorité de sujets, une implication inverse pour quelques autres, des « *profils* » correspondant à des « *styles* » différents d'expression, voire une certaine indépendance d'ensemble...

Se pose alors, une question plus générale, de l'ordre pratiquement épistémologique, sur le cadre même des conceptions de la recherche. Supposons que l'on puisse réduire le sujet à un *système*, ce qui est déjà discutable, est-il pertinent de traiter de dimensions dépendantes des « *systèmes* »/sujets autonomes, dynamiques, *autopoïétiques*, indépendamment de l'organisation globale de ces *systèmes* ? Et jusqu'où cela est-il pertinent de le faire ? Surtout, est-il pertinent d'en extraire des mesures sur des dimensions artificiellement isolées pour en inférer des effets de ces dimensions isolées les unes sur les autres ? Les réponses (normandes ?) à ces questions peuvent être oui et non. Non si un minimum de maîtrise du contexte dans lequel les *systèmes* évoluent ne sont pas connus suffisamment pour saisir les interactions susceptibles de fonctionner entre les dimensions considérées. Non si la connaissance du *système*, ou de sous ensembles bien définis de ce *système*, ne sont pas connus. Oui, sans doute, si les conditions sont suffisamment définies. Oui, notamment si le chercheur vise à vérifier le bien fondé d'un modèle même schématique de compréhension *locale* du fonctionnement du *système* considéré, ou de l'interaction fonctionnelle, elle-même potentiellement systémique, de ce système, ou d'un sous-système, avec ses contextes. Oui, également, si le chercheur émet a priori un ensemble d'hypothèses susceptibles d'être clairement confirmées ou contredites par les résultats de la recherche. L'analyse implicative peut donc être utile dans ce sens aux conditions rigoureuses de l'analyse de la qualité du cadre de recherche pour lequel elle est convoquée. Une qualité première est la clarté des choix faits et donc des risques assumés par le chercheur. Ceci étant dit, sur ce point, deux positions globales différentes, au plan épistémologique, peuvent être prises au regard de la figure complexe que représente la répartition des données sur le tableau ci-dessus.

Une position consiste à se demander si l'hypothèse d'une logique générale d'implication peut être retenue. Si le chercheur décide que la réponse est oui, alors il se trouve dans la position de considérer que les exceptions à la règle sont négligeables et qu'elles peuvent tenir à de multiples facteurs ne remettant pas en cause le principe logique retenu de l'implication.

Une autre position, à l'opposé, consisterait, plutôt, à ne voir, à l'extrême, que des sujets référant à des situations, des contextes singuliers. Ces contextes, systématiquement différents pourraient être présents en arrière plan de *chaque croisement de données*. Certains ne seraient pas visibles et les profils seraient identiques parce que les différences ne tiendraient pas, par exemple, aux deux dimensions ici retenues. Elles n'apparaîtraient que si d'autres dimensions étaient associées. Certaines des différences apparaîtraient, cependant, ici, sous formes de profils de scores différents. Il y aurait des profils différents que l'on pourrait regrouper, en termes de similitude, *dans le contexte considéré*, selon une logique traduisant une *co-*

*variation* de scores sur les deux dimensions retenues, à côté de profils *asymétriques*. Concernant les profils asymétriques, différentes configurations regroupant des formes, là aussi comparables, de sujets différents, peuvent se traduire par différentes interprétations. Ces interprétations peuvent être, comme cela a déjà été dit, de l'ordre de *l'incompatibilité* entre les dimensions (traduit par une corrélation négative pour une partie des profils asymétriques). Elles peuvent être de l'ordre d'une *implication* possible dans un sens, pour une part des données, de *l'implication en sens inverse*, pour une autre part, ou, tout simplement, d'une *typologie de profils différents* d'un sujet à une autre.

Mais une troisième position, « *mixte* », peut exister. Elle passe nécessairement par des considérations en termes de *contraintes* exercés par les contextes internes ou externes agissant de manière plus ou moins large (générale) et surtout *identique*, sur des caractéristiques de sujets, considérés, eux, comme différents. Cette position *mixte* consiste à reconnaître une forme dominante relevant d'une part importante de l'information contenue dans la forme de distribution croisée des données dont peut rendre compte l'analyse statistique. Mais cette reconnaissance d'une part dominante d'ordonnement de l'information ne nierait pas, pour autant, l'expression, possible ou effective, de formes conjointes différentes traitant de parts moins importantes de l'information observée, voire la reconnaissance d'un profil singulier d'un sujet. Ces « *profils* » individuels ou de sous-groupes, seraient porteurs de sens dans la recherche parce qu'ils traduiraient des expressions originales inattendues (sérendipité) source d'ouverture à des parts de liberté dans un jeu de contraintes plus globales. Les différences de « *poinds* » des différentes formes d'expression traduiraient simplement des contraintes (toujours partielles ?) plus ou moins fortes et des plus ou moins grandes parts de liberté possibles dans ce jeu de contraintes imparfaites d'un système complexe (Pellois, 2005, 2008, 2010). Ces parts de libertés possibles, plus ou moins connues, constitueraient des ressources ou potentialité pour « *faire autrement* », en soi, ou pour aboutir aux mêmes habiletés mais par d'autres profils de scores (vicariance). La question reste cependant posée de comment dissocier l'erreur de mesure de l'expression d'une position originale *effective*. Cette question pourrait être traitée par le biais d'une approche individuelle de type « *étude de cas* », prenant, notamment, en compte d'autres informations concernant le cas étudié.

Pour conclure provisoirement sur ce point, se trouve être posé, ici, du général (population globale) à l'individuel, en passant par la position différentielle (sous-populations), un cadre méthodologique potentiel d'interprétation et d'analyse des données, cadre dans laquelle le logiciel CHIC, comme outil de traitement, sans doute, à côté d'autres, peut prendre tout son sens (Cf. Chamsine, 2012, à paraître).

#### **4 Traitement par l'analyse implicative de ces différentes formes d'ordonnement de données**

Le logiciel CHIC, propose, d'une part, de traiter des données en termes de similarité, d'implication et, sous une forme qui tient compte des deux précédentes, de cohésion. D'autre part, le logiciel propose différentes options de calcul (entropie/classique, loi de Poisson/binomial). Qu'en est-il, en termes de résultats de l'analyse statistique effectuée par le logiciel CHIC, de ces différentes distributions

théoriques de données des tableaux du point 3 ci-dessus ? Et, en fonction des résultats trouvés, existe-t-il des indications qui alerteraient sur des particularités de formes de distribution des données recueillies ? L'objectif de ce point 4 est donc de soumettre au calcul les distributions de données précédentes, puis de confronter les formes modélisées, aux résultats fournis par le logiciel CHIC, ceci afin d'en tirer, éventuellement, quelques enseignements.

#### 4.1 Un tableau d'observation simple d'implication stricte

Le premier tableau soumis au traitement statistique par CHIC est le tableau simplifié d'observation de la fin du point 2.1. (Cf. Tableau 1), présentant des observations répondant strictement à l'implication logique et dans lequel les positions « *lampe allumée* » et « *lampe éteinte* » ont été affectées de la valeur 0 ou 1, ceci pour permettre le traitement par le logiciel.

Les résultats sont, bien sûr, ceux que l'on pouvait attendre dans une telle situation. Pour toute les formes de traitement l'indice d'implication  $A \Rightarrow B$  est de 100, et ceci quelque soit la modalité de traitement retenue. Par ailleurs l'indice de similarité est, lui, de .99. Une seule différence tient au passage du traitement sur le mode « *classique* » au traitement sur le mode « *entropique* ». Lorsque le traitement par CHIC est effectué sur le mode « *classique* » l'indice conjoint à  $A \Rightarrow B$  (logique Implicative réciproque :  $B \Rightarrow A$ ) est de 98 (sensiblement de même valeur que l'indice de similarité), cette valeur est 0 lorsque le traitement est sur le mode « *entropique* » et ceci quelque soit la loi du traitement statistique utilisée par ailleurs (loi de Poisson ou binomiale). Ce point de différence est le seul observé ici. Comme ces différentes modalités statistiques sont préconisées en fonction de la taille de l'effectif des observations faites, il sera simplement utilisé, par la suite, pour un effectif systématique de cent observations modélisées, un traitement sur le mode « *classique* » avec application de la loi binomiale<sup>15</sup>. Selon les informations fournies par le concepteur du modèle d'analyse statistique, le traitement sur le mode « *entropique* » serait plutôt réservé à des observations en grand nombre (quelques centaines, des milliers) et l'utilisation de la « *loi de Poisson* » serait plutôt réservée à un nombre d'observations qui serait, au plus, de quelques dizaines.

#### 4.2 Un tableau d'observation plus complexe présenté sous la même forme globale que celui de l'« *implication stricte* »

Lorsque l'on passe à un tableau de performances s'inscrivant selon un continuum croissant et lorsque que sont reprises des répartitions similaires sur la base de 4 secteurs de type (plutôt 0,1 ou 1,0, ou 1,1, ou 0,0 ; Cf. Tableau 2) il est clair que, comme évoqué à la section 2.2, si la majorité des observations faites vont dans le sens d'une tendance implicative (valeurs de B supérieures à A), des observations (valeurs de A supérieures à B) infirment la tendance implicative majoritaire. Les résultats du traitement nous donnent un indice de similarité de .64. Cet indice modérément élevé s'interprète assez facilement puisque nous avons un nombre d'observations conjointes (A/B) placées dans

---

<sup>15</sup> Des essais de traitements conjoints ont montrés que pour cent observations les différences quantitatives des résultats obtenus par traitement sur le mode « *loi de Poisson* » et le mode binomial étaient relativement faibles.

le secteur inférieur droit du tableau (position 0, 1) élevé (50, soit la moitié des observations). L'indice d'implication, lui est de 78 dans le sens attendu ( $A \Rightarrow B$ ). Les données du tableau correspondant aux indices d'implication sont nettement asymétriques puisque l'autre indice est de 65. Les indications en termes d'indice de cohésion vont, bien entendu, dans le même sens (.65 contre .36).

### **4.3 Un tableau d'observation plus complexe mais dont la répartition des données respecte le principe de la *tendance implicative* dans toutes les situations considérées**

Qu'en est-il des résultats lorsque toutes les observations respectent la tendance implicative de type  $A \Rightarrow B$  ? Pour cela toutes valeurs doivent être dans une position d'asymétrie telle que, pour chaque score observé en A, lui correspond un score observé en B toujours supérieur. C'est le cas de figure du tableau 3 présentés au point 2.2.1.

Les résultats du traitement par CHIC, indiquent que, bien qu'aucune valeur ne se trouve dans la diagonale du tableau l'indice de similarité (.78) est supérieur au précédent. Ce résultat est cependant conforme à une observation plus détaillée qui permet de s'apercevoir que la majorité des valeurs du tableau, bien que respectant la logique asymétrique, sont voisines, ceci à un (par exemple : 0,2 et 0,3) ou deux degrés (par exemple : 0,2 et 0,4) de différence de score près (52 observations conjointes) et que les observations à plus de 4 degrés de différence de score (par exemple : à partir de 0,2 et 0,7) ne représente qu'un nombre d'observations de l'ordre de moins d'un quart (23) de l'ensemble des observations conjointes de A et de B.

L'indice d'implication, lui est de 98 largement supérieur, comme attendu, à l'indice du tableau précédent. Il va, bien sûr, également dans le sens attendu ( $A \Rightarrow B$ ). Les données du tableau correspondant aux indices d'implication sont, là aussi, nettement asymétriques puisque l'indice inverse est de 79. Les indications en termes d'indice de cohésion vont dans le même sens (.99 contre .67).

### **4.4 Un tableau dont la répartition des données respecte très majoritairement le principe de la *tendance implicative* avec quelques situations contraires à cette tendance**

Le tableau considéré ici est le tableau 4 du point 2.2.2, pour lequel il existe aussi une tendance à une co-variation, sensiblement similaire au Tableau 3, quoique nettement plus accentuée (présence de valeurs dans la diagonale du tableau).

Les résultats fournis par CHIC donnent bien un indice de similarité (.79) supérieur aux deux précédents. L'indice d'implication va bien dans le sens attendu avec une valeur légèrement inférieure à la précédente (96) mais supérieure à celle du point 4.2. L'asymétrie avec l'indice inverse (82) est également plus réduite. Les indications fournies concernant l'indice de cohésion vont dans le même sens (.97 contre .74).

## 4.5 Des tableaux traitant de répartitions des données référant à la corrélation voire à une forme d'« incompatibilité »

### 4.5.1 Corrélation entre les deux séries de données

Le tableau considéré ici est le tableau 5 de la section 3.1. Comme attendu, le coefficient de corrélation comme l'indice de similarité sont élevés (.83 et .74). Les indices d'implication et de cohésion également, mais comme attendu, leurs valeurs A/B et B/A sont équivalentes (symétrie des données). Elles sont de 86 pour l'implication et de .74 pour la cohésion. De manière surprenante, malgré ces valeurs symétriques, le graphe, au seuil de 75 pour l'implication, donne une indication d'implication dans le sens  $B \Rightarrow A$  et une indication de cohésion dans l'autre sens... Enfin il est à noter que la corrélation est forte, comme l'indice de similarité (cela paraît cohérent), mais aussi, que ceux de l'implication et de la cohésion, certes équivalents dans les deux sens, sont élevés.

### 4.5.2 Corrélation négative entre les séries de données : notions « d'incompatibilité », voire de « contre-indication » ?

Il s'agit de traiter ici du tableau 6 de cette même section 3.1. Comme la répartition des données, ici, est la même que précédemment concernant la corrélation, mais en position de corrélation négative le coefficient de corrélation est le même mais affecté du signe négatif (-.83). L'indice de similarité est inférieur à .50 (.26). Les indices d'implication ont la valeur 14 tous les deux, et ne sont donc pas pris en compte sur le plan de la représentation graphique puisqu'inférieurs à .50. Les indices de cohésion, eux, sont nuls.

## 4.6 Des tableaux traitant de répartitions asymétriques des données pouvant relever d'interprétations en termes de « facile » « difficile » ou de la notion d'effet « seuil »

### 4.6.1 Distributions pouvant référer à la notion de « difficile » pour une des épreuves considérées

Il s'agit, ici, du tableau 7 présenté au point 3.2. Rappelons que dans ce tableau, peu de valeurs de B dépassent le score de 0,5 (5 seulement) alors que les valeurs de A se répartissent équitablement par tranches de scores identiques (effectif 10 pour 0,0, 10 pour 0,1, etc.). Les résultats obtenus donnent un coefficient de corrélation de .32 et un indice de similarité de .70. Les indices d'implication et de cohésion sont élevés et nettement asymétriques. Ils caractérisent une implication  $B \Rightarrow A$  (indice de 83 contre un indice traitant de l'implication inverse de 65) et une cohésion allant dans le même sens (.79 contre .39, le nœud de niveau 1 ayant une valeur significative).

### 4.6.2 Distributions pouvant référer aux notions de « difficile » « facile » pour les épreuves considérées

Il s'agit de traiter ici du tableau 8 présenté au point 3.3. Les valeurs résultant du traitement par CHIC sont, ici, très contrastées par rapport aux valeurs obtenues dans le cas précédent. En effet la corrélation est négative (-.15). L'indice de similarité est

inférieur à .50 (.45). Les indices d'implication sont également inférieurs à 50 et asymétriques (47 dans le sens A/B contre 32 dans le sens B/A). Les indices de cohésion sont, eux, nuls.

#### 4.6.3 Distributions pouvant référer à une « logique » de « seuil »

Il s'agit de traiter ici du Tableau 9 présenté au point 3.4. Dans cette logique de « seuil », l'essentiel des valeurs de la variable B sont au-delà de 0,5 à l'exception de deux valeurs. Les valeurs de A sont plus nombreuses au-delà de 0,5, également, mais elles couvrent l'ensemble de l'échelle de valeurs. Leur nombre, en deçà de 0,5 est cependant non négligeable (26 valeurs). Ceci fait que ce tableau initialement construit pour traiter de la question de seuil pourrait tout aussi bien traiter du rapport entre un indice relevant d'une logique réfèrent à la notion de « facile » et d'un indice traitant d'une logique réfèrent à la notion voisine « un peu moins facile ». La corrélation fournie par CHIC est de .62, l'indice de similarité est de .61. Les indices d'implication et de cohésion sont nettement asymétriques. Les résultats obtenus valident une tendance implicative de type  $A \Rightarrow B$  (85 contre 72) et une indication en termes de cohésion allant dans le même sens (.79 contre .51). Les valeurs de similarité et de cohésion amène le traitement statistique par CHIC à considérer le nœud de niveau 1 comme significatif dans les deux cas.

### 4.7 Des tableaux traitant de situations plus complexes

#### 4.7.1 Distribution référant à une notion d'« optimalité »

Il s'agit de traiter ici du tableau 10 de la section 3.5. Rappelons que cette distribution de données présente une évolution croissante des valeurs de A pour une première partie de la croissance des valeurs de B (en gros, du minimum à 1/3 de sa valeur), puis une constante des valeurs de A, au maximum du score, pour une deuxième partie de la croissance des valeurs de B (partie médiane de la croissance des valeurs de B), enfin une décroissance des valeurs de A pour la dernière partie de la croissance des valeurs de B vers le maximum de sa valeur. Les résultats obtenus après traitement par CHIC sont les suivants. Le coefficient de corrélation est nul mais affecté d'une indication négative (-0,00). L'indice de similarité est de .50. Les indices d'implication, de 50, sont équivalents. Les indices de cohésion sont nuls.

Deux groupes sont dissociés dans des analyses statistiques complémentaires, l'un pour lequel la valeur de A croit avec la partie correspondante de la croissance de B, soit dans des proportions équivalentes au groupe inverse, soit dans des proportions respectivement de l'ordre de 2/3, et 1/3 pour le groupe inverse.

Comme attendu, pour ces groupes, relevant de la logique croissance de B/croissance de A, les corrélations sont positives et élevées (.91 et .83), pour les groupes, relevant d'une logique inverse de croissance de B/décroissance de A elles sont négatives (-.91 et -.94).

Pour les groupes relevant d'une croissance associée des scores de A et B, les indices de similarité sont élevés (.72, .77). Les indices d'implication sont également élevés et nettement asymétriques suggérant *une tendance implicative forte de type  $B \Rightarrow A$*  (98 contre 67 et 99 contre 76) *liée à la forme de répartition des fréquences statistiques des*

*données observées mais inverse d'une possible implication logique réelle liée à la compréhension de l'action d'une des dimensions sur l'autre* (Cf. section 3.5), *possible implication renvoyant à un autre tableau ordonnant autrement les résultats obtenus*. Il en va de même concernant la valeur et le sens des indices de cohésion (.99 contre .44 et 1.00 contre .60).

Pour les groupes relevant de la tendance inverse, l'indice de similarité est inférieur à .50 et les indices d'implication sont inférieurs à 50. Les indices de cohésion, eux, sont nuls.

#### **4.7.2 Distribution pouvant référer à une double « logique » implicative différentielle**

D'autres formes plus complexes de distribution de données peuvent apparaître suggérant d'autres logiques que celles-ci-dessus évoquées, c'est le cas, par exemple, du tableau 12 présenté à la section 3.6. Comme attendu, vu la forme générale de distribution des données, la valeur de la corrélation est élevée (.88), de même que l'indice de similarité (.93, présenté comme non significatif par le logiciel). Les indices d'implication sont également très élevés mais très faiblement asymétriques (99 contre 98) indiquant la présence d'une tendance implicative légèrement asymétrique de type  $B \Rightarrow A$ . Ce résultat est, bien sûr, conforme au résultat attendu du fait que la tendance implicative B/A a été volontairement recherchée par un poids des fréquences de distribution de données en ce sens légèrement supérieur aux fréquences de distribution de données favorables à la tendance implicative inverse (A/B). La tendance cohésive va dans le même sens mais les indices cohésifs sont quasiment identiques (ils sont affichés tous les deux à .99). Cependant le graphe indique tout de même une tendance de type B/A).

Dans la mesure où la forme de distribution de données a été construite pour simuler deux sous-formes traitant de logiques implicatives inverses en fonction des scores obtenus sur une des dimensions (A), il a été procédé à la constitution de deux sous-groupes : l'un avec des scores de A allant de 0,5 à 0,9, traitant de la forme d'implication conforme au résultat global obtenu, l'autre avec des scores de A allant de 0,0 à 0,4, traitant de la forme d'implication inverse. En ce qui concerne le premier sous-groupe, la corrélation est de .51 mais l'indice de similarité paraît faible (.53) au regard de l'indice du groupe global et il n'est pas significatif au sens du logiciel. Les indices d'implication, comme pour le groupe global, vont bien dans le sens attendu ( $B \Rightarrow A$ ) avec une asymétrie claire entre les deux indices (63 contre 58). Il en va de même pour les indices de cohésion mais les valeurs sont nettement plus faibles (.32 contre .20) que celles du groupe global. La liaison est cependant affichée comme significative. En ce qui concerne le second sous-groupe, la corrélation est de .54. L'indice de similarité, quoique plus élevé que celui du premier sous-groupe, est, encore, beaucoup plus faible (.63) que celui du groupe global. Cependant, il est significatif pour le logiciel. Les indices d'implication vont bien dans le sens attendu (cette fois-ci :  $A \Rightarrow B$ ), inverse du groupe global, avec également une asymétrie claire entre les deux indices (67 contre 60). Il en va de même pour les indices de cohésion mais avec des valeurs, là aussi, nettement plus faibles (.40 contre .25) que celles du groupe global. Elles sont, cependant, ici aussi, significatives.

## 4.8 Autres tableaux référant à des situations différentielles ; vers une logique d'individualisation de l'analyse de formes complexes ?

### 4.8.1 Distribution de données et *vicariance*

Ce point traite du tableau 13 présenté à la section 3.7. La corrélation a une valeur négative très élevée (-.92). L'indice de similarité est très faible (.07). Le nœud au niveau 1 est néanmoins indiqué comme significatif. Les indices d'implication sont identiques. Les indices de cohésion sont nuls.

De la même façon que précédemment, et selon les mêmes modalités, deux sous-groupes ont été générés. Les deux tableaux correspondant à ces sous-groupes présentent les mêmes corrélations (-.36), les mêmes indices de similarité (.48), les mêmes indices d'implication asymétriques (48, 37) mais les logiques implicatives sont inversés d'un sous-groupe à l'autre. Les indices de cohésion sont nuls pour les deux sous-groupes.

### 4.8.2 Distribution pouvant référer à des logiques individuelles plurielles

Il s'agit de traiter ici du tableau 14 présenté à la section 3.8. La corrélation obtenue entre les scores référents aux deux dimensions est de .17. L'indice de similarité est de .58. La liaison n'est pas considérée comme significative par le logiciel. Les indices d'implication font valoir une tendance implicative de type  $B \Rightarrow A$  avec une asymétrie affichée (65 contre 61). Les indices cohésifs, quoique présentant des valeurs plus faibles (.37 contre .26), vont dans le même sens. La liaison est néanmoins affichée comme significative.

Une dissociation en deux sous-groupes (premier sous-groupe avec des valeurs de B comprises entre 0,5 et 0,9, et un second sous groupe avec des valeurs allant de 0,0 à 0,4) a aussi été effectuée ici.

En ce qui concerne le premier sous-groupe la corrélation est de .18, l'indice de similarité est de .52 (valeur significative pour le logiciel), la tendance implicative de type  $A \Rightarrow B$  apparaît (55 contre 54), les indices de cohésion vont dans le même sens (.12 contre .10). Là aussi la tendance est affichée comme significative. En ce qui concerne le second sous-groupe la corrélation est négative (-.12), l'indice de similarité est de .47 (non significatif), les indices d'implication de type  $B/A$  sont de 47 contre 43, inférieurs à .50. Les indices de cohésion, eux, sont nuls.

## 5 Au regard des résultats obtenus : quelques remarques et quelques points de discussion

### 5.1.1 Variation des formes de distribution de données et application du logiciel CHIC

Il n'a été présenté ici, par un souci évident de concision, qu'un nombre limité de cas de figures des différentes formes de distribution de données susceptibles d'exister. Une multitude d'autres formes pourrait être envisagée. A l'issue de ce travail exploratoire, il est clair qu'il existe des rapports complexes entre variation des formes et traduction,

notamment en termes d'implication, lorsque le logiciel CHIC est appliqué. De manière synthétique, qu'en est-il des observations effectuées au cours de ce travail ?

a) Lorsque les données se répartissent selon une distribution simple type tableau à quatre cases, les résultats apparaissent clairement et sans ambiguïté (Cf. section 4.1, Tableau 1). Ils montrent cependant des particularités qui mériteraient quelques explications concernant le logiciel et son utilisation. Ceci permettrait notamment de clarifier l'utilisation qui peut être faite des modes « *classique* » et « *entropique* », du traitement statistique, ces explications ne se limitant pas nécessairement à une question de taille d'échantillon des données traitées.

b) La question du rapport entre analyse implicative et analyse cohésive, qui peuvent aboutir parfois, à la marge, à des résultats déroutants voire contradictoires (graphes inversés l'un par rapport à l'autre, Cf. section 4.5.1, en référence au tableau 5), aurait, sans doute, besoin d'être mieux explicitée ceci en vue d'une meilleure interprétation conjointe des résultats obtenus.

c) Des aspects de l'analyse implicative par CHIC peuvent être saisis de manière assez directe, par exemple : l'implication dans un sens donné entre deux dimensions quantifiée par deux indicateurs asymétriques (Cf. sections 4.2, 4.3, Tableau 2 (78/65) et 3 (98/79)) ; d'autres le sont moins facilement, par exemple : il apparaît souvent, dans les résultats, des indices de similarité voire des coefficients de corrélation, élevés, plus ou moins, associés à des indices d'implication élevés<sup>16</sup>. Qu'en est-il du rapport, dans CHIC, entre ces deux catégories d'information statistique ? De même, outre le fait qu'aux corrélations positives ou négatives correspondent des niveaux de l'indice de similarité supérieur ou inférieur à .50, qu'en est-il exactement du rapport entre coefficients de corrélation et indices de similarité (Cf. les variations constatées dans la note de bas de page) ? Il semble exister une relative indépendance de variation entre les deux non nécessairement liées à la taille du groupe considéré. En ce qui concerne le tableau 2 (section 4.2), et le tableau 3 section 4.3), les coefficients de corrélation et les indices de similarité sont respectivement de .23 et .64 dans le premier cas pour .64 et .78 dans le second cas, ceci pour des effectifs équivalents d'observations conjointes.

d) Lorsque la forme de distribution de données relève de la quasi implication, le chercheur peut estimer qu'une proportion de 4%<sup>17</sup> « *d'erreur* » peut constituer une bonne estimation d'une tendance implicative avérée. Les résultats d'une telle situation apparaissent en section 4.4, tableau 4. L'indice d'implication, dans le sens attendu (A=>B), est de 96. Comment interpréter, dans ce cas de figure, l'indice de la contraposée de 82 ? L'écart peut-il être plus élevé avec 4% de cas contraires dans une autre répartition ? Si oui, dans quelle forme de configuration des données cela peut-il se faire ? Par accentuation des contrastes des scores allant dans le sens de l'implication ? Par accentuation du contraste pour les cas inverses ? Ces deux raisons associées ? D'autres encore ?

e) Les résultats concernant les situations de corrélation positive et négative (« *incompatibilité* » ?) méritent également d'être commentés (Cf. section 4.5.1, Tableau

<sup>16</sup> Dans le cas d'un mode de traitement « *classique* » et selon la loi binomiale. Exemple, pour les Tableaux 2, 3, 4, 5 : coefficients de corrélation respectivement de .23, .64, .70, .83, indices de similarité .64, .78, .79, .74, indices d'implication 65/78, 79/98, 82/96, 86/86.

<sup>17</sup> Donc ici 4 cas d'observations sur 100, observations également à la marge de l'ensemble des mesures retenues.

5 et section 4.5.2, Tableau 6). Les résultats concernant ces deux répartitions vont dans le sens attendu avec : des corrélations positives et négatives exactement symétriques (.83 et -.83), des indices de similarité dont le premier est élevé (.73) et le second est inférieur à .50 (.26), des indices d'implication et de cohésion élevés dans le premier cas (.86/.86, .81/.81), faibles ou nuls dans le second. Mais, s'il y a symétrie, et donc si l'asymétrie (du fait de la construction de la répartition des données) n'existe pas au plan quantitatif, les indices d'implication sont cependant élevés dans le cas de la corrélation positive et le logiciel se prononce en terme de graphe sur la liaison (B/A pour l'implication, à 75, et A/B pour la cohésion ... la distribution des valeurs étant, cependant, par construction, strictement symétrique). Ceci peut être gênant dans une interprétation hâtive de résultats de recherche, formulée dans une confiance absolue et sans discussion, après observation détaillée des résultats obtenus.

f) Rappelons, que des distributions asymétriques de données (fréquences élevées vers les scores faibles ou élevés) posent des contraintes qui font problèmes en termes d'interprétation et de sens (Cf. notions d'épreuve « facile », « difficile », « adaptée »). Dans le cas de figure d'épreuves « facile »/« adaptée », référant au tableau 7, à côté d'un indice de similarité de .70, apparaît, après calcul par CHIC, une tendance implicative de type  $B \Rightarrow A$ , bien caractérisée (indice 83 contre 67). Il en va de même concernant l'indice de cohésion (.76 contre .39). Du point de vue strictement de l'observation de la distribution des données ce résultat est conforme à l'hypothèse implicative, mais c'est cette hypothèse, justement, qui pose ici problème au regard d'un autre sens possible qu'aurait la forme de distribution des données : celui de l'inégale difficulté, pour les sujets considérés, des épreuves proposées. Cette inégale difficulté ne justifie pas nécessairement, en termes de sens, l'hypothèse implicative, ceci même et y compris si le logiciel confirme par ses calculs, du fait de l'ordonnement des données recueillies, cette hypothèse.

Dans un autres cas de figure de même type, les observations renvoient au cas de figure correspondant à des épreuves « facile »/« difficile » (Cf. Tableau 8). Les résultats obtenus, après traitement des données par CHIC, ne traduisent pas, comme dans le cas précédent, une « tendance implicative » dont l'interprétation serait discutable. Cela semble tenir au fait que la corrélation étant négative, les indices de similarité et d'implication sont inférieurs à .50.

g) Un autre type de distribution de données illustre la difficulté que peut revêtir la recherche de sens et son rapport aux résultats fournis par le logiciel. Ce type de distribution réfère au tableau 9. La distribution des données de ce tableau traitant de l'effet de « seuil » peut renvoyer tout aussi bien à une logique d'épreuves « facile/facile, plus ou moins ». Les résultats obtenus, après traitement des données par CHIC, comme en ce qui concerne le tableau 7, donnent des indices caractérisant une tendance implicative de type, cette fois-ci,  $A \Rightarrow B$  (indice 85 contre 72). Il en va de même concernant les indices de cohésion (.79 contre .51). Notons que la logique implicative correspond, cependant, ici, à quelque chose de pertinent, du point de vue d'un « effet seuil ». En effet, l'effet seuil suppose, dans le cas de figure considéré, que l'expression de A se fait **sous la dépendance** de B mais ceci à partir d'un certain seuil (ici situé autour du score de 0,5). Cette logique de « dépendance » serait donc bien en adéquation, sous condition particulière, avec l'implication de type  $A \Rightarrow B$ .

h) L'exemple traité à la section 4.7.1 (Tableau 10) montre, *a contrario*, que, pour un autre contexte de distribution de données, cette convergence entre logique implicative relevant du cadre de recherche et résultats du traitement par CHIC, conformes à cette logique, n'existe pas. Les résultats obtenus peuvent même afficher une tendance qui sera considérée comme « *erronée* », voire « *contraire* » au regard du contexte particulier de la forme de distribution, ici celle d'une relation d'« *optimalité* ». Ce n'est bien évidemment pas le logiciel qui est en cause mais la qualité du cadre du raisonnement utilisé. Cette « *non qualité* » relève, ici, de la forme de présentation de la distribution des données choisies, forme qui sera proposée au traitement par CHIC (Cf. Tableau 11).

i) Des indices d'implication peuvent être quantitativement proches (Cf. le point 4.7.2, et les résultats concernant le Tableau 12). Ce peut être le signe de distributions complexes de données. Par exemple, cela peut être le signe, comme c'est le cas ici, d'une double tendance s'exprimant en fonction des scores sur une des deux dimensions considérées. Cette double tendance amène, alors, à un changement de point de vue qui ferait passer d'un cadre (conceptuel, méthodologique) généraliste à un cadre différentiel et amènerait à y associer les analyses statistiques adaptées. Cette approche différentielle « *en aval* », passe, après constat de la proximité des indices implicatifs et observation concrète de la distribution des données, par la constitution de deux sous groupes permettant de vérifier, après analyse par le logiciel CHIC, la validité de l'hypothèse concernant la tendance implicative directe ou inverse en fonction du sous groupe considéré. Mais, suite à différents essais sur des distributions de données différentes, il semblerait que cette approche ne soit pertinente que dans le cas de dimensions, par ailleurs corrélées entre elles.

j) Plus radicalement, la forme de distribution de données peut relever d'autres interprétations renvoyant à des différences d'expression en termes de « *profils* » (performances cognitives, traits de personnalité) pour lesquels la question de la relation entre les dimensions considérées (co-variation, *incompatibilité*, implication, etc.) ne se pose pas nécessairement, ou, si elle se pose, cela se fait en terme de vicariance, de cheminements alternatifs, par exemple dans le cadre de l'étude du développement psychologique différentiel de l'enfant. Des formes peuvent traduire ces différences ou ces cheminements vicariants. Pour autant elles ne se traduisent pas, en tant que tel, en termes de résultats obtenus par leur traitement par CHIC. Cela peut tenir au fait qu'elles ne relèvent pas particulièrement du cadre d'analyse du logiciel. C'est le cas (Cf. Tableau 5) lorsque ces formes renvoient, par exemple, à des corrélations négatives. Mais quand, de plus, la forme peut renvoyer à des comportements suggérant l'aspect différentiel (Cf. Tableau 13), il conviendrait, avant d'effectuer le traitement par CHIC, d'inscrire d'emblée, dans la conception même de la recherche, cette dimension différentielle. Cette conception différentielle pourrait retenir des groupes différents en termes de profils établis sur les deux dimensions considérées. La dimension dominante de chaque typologie de profil serait alors confrontée à un troisième indicateur sur lequel l'action *vicariante* est postulée, par exemple, le développement cognitif. L'application de CHIC serait alors pertinente pour chacun des groupes considérés et non pas pour les deux dimensions initiales et le grand groupe dans son ensemble. Le cadre différentiel (méthodologique, conceptuel) ne serait pas posé, ici, en « *aval* », comme dans le cas précédent, mais en « *amont* ».

k) Des formes plus complexes, plus atypiques encore peuvent apparaître. Qu'en est-il de la pertinence de l'utilisation du logiciel CHIC dans le cas de figure d'une forme de distribution de données, par exemple du type de celle du tableau 14 ? Il est clair que dans ce cas de figure le calcul par CHIC aboutit à faire apparaître la tendance dominante, à savoir, la co-variation/implication associée à une certaine non dispersion des données, par un indice de similarité sensiblement supérieur à .50 et un indice d'implication légèrement asymétrique dans le sens  $B \Rightarrow A$ , les indices de cohésions étant, eux, inférieurs à .50. Mais la diversité des situations d'observation conjointe, tel qu'elle apparaît concrètement à l'examen du tableau, n'apparaît bien évidemment pas clairement dans les résultats obtenus. Mais, en l'absence de vision « clinique » de la forme de distribution des données, il semble raisonnable de penser que les résultats globaux obtenus par CHIC devraient alerter, là encore, le chercheur sur une possible complexité qui échappe à leur seule lecture. Le chercheur devrait alors changer de forme d'approche des données qu'il a en sa possession et passer d'une analyse statistique globale à des analyses différentielles fines, voire au traitement individuel (étude de cas ?) de chaque situation observée...

Ces différents cas de figure, ci-dessus évoqués, montre bien la nécessité d'analyser, discuter, interpréter les données hors du traitement informatique par le logiciel, voire de traiter, en amont de la méthodologie de recherche et de son rapport, en termes d'adéquation, au mode de traitement (ici statistique) utilisé, c'est ce dont il va être question maintenant.

### 5.1.2 Des formes complexes de distribution de données à la question des contraintes de contexte, du différentiel et de l'approche individuelle

Il semble clair que de l'implication prise au sens logique strict à la « *tendance quasi implicative* » mesurée sur un ensemble de faits observés, il y a un passage à un niveau de complexité méthodologique qui nécessite une réflexion d'ensemble portant sur la pratique d'utilisation de l'analyse implicative en général, du logiciel CHIC en particulier. Peuvent être distingués ici, en première approximation, deux niveaux donnant lieu à plus ou moins de précautions dans la réflexion de la part du chercheur.

a) Lorsque les analyses prennent des formes simples, par exemple, sur des données dichotomisées, les risques d'erreur d'interprétation paraissent faibles. La fiabilité des résultats directs obtenus peut être considérée comme bonne quant aux interprétations qui peuvent être avancées et ceci d'autant plus que les résultats obtenus sont très contrastés.

b) Lorsque l'analyse porte sur des données plus complexes, par exemple des échelles de mesures s'inscrivant sur un continuum, surtout lorsque les résultats sont faiblement contrastés, la situation paraît moins évidente. Aussi le chercheur devrait envisager deux perspectives.

- L'une consisterait à examiner concrètement et plus directement la tendance que peut prendre la forme de distribution des données et déterminer le sens qu'elle peut avoir.

- l'autre concernerait l'estimation des contraintes, notamment contextuelles, qui peuvent peser globalement et l'opportunité qu'il y aurait, éventuellement, à mettre en œuvre une approche différentielle (voire individuelle) des observations recueillies. Dans

ce cas il convient d'éviter un renvoi trop systématique mais commode à des erreurs de mesure, certes toujours possibles.

## 6 Conclusion : pour une qualité de la recherche utilisant l'analyse implicative, quelles règles à mettre en œuvre ?

Il paraît souhaitable, cependant d'aller plus loin sur cette question de la qualité de la recherche. En ce sens, comme pour d'autres outils d'analyse de données, les recherches utilisant CHIC doivent respecter des règles, pour certaines déjà bien connues du chercheur<sup>18</sup>, ceci d'autant plus que l'outil de traitement de données utilisé est sophistiqué. Une des difficultés tient, ici, entre autre, à la particularité du mode de traitement asymétrique des données de l'analyse implicative.

Quelques unes des règles qualitatives d'utilisation du traitement de données, plus spécialement par CHIC, peuvent donc être évoquées, en conclusion de ce texte, même si cette liste est, sans doute, et encore à ce stade de la réflexion, bien incomplète.

Il conviendrait, pour le chercheur utilisant l'analyse implicative :

- de disposer d'un modèle structuré, formellement cohérent, donnant lieu à des hypothèses précises et susceptibles d'être clairement infirmées ;
- de développer un traitement « *pas à pas* », au moins dans un premier temps, ceci afin de maîtriser les liaisons deux à deux des variables avant d'engager des analyses plus « *structurales* » ;
- de traiter de l'analyse « *structurale* » plutôt de manière **confirmatoire**, notamment au regard du modèle « *structural* » hypothétique initialement développé, sans pour autant rejeter des aspects **exploratoires** notamment au regard de résultats non attendus (*sérendipité*), ne serait-ce que pour ouvrir à de nouvelles questions voire permettre la formulation de nouvelles hypothèses, de nouvelles modélisations, qu'il conviendrait de soumettre à l'expérimentation dans une étape ultérieure des recherches ;
- de confronter les analyses structurales aux résultats des analyses entre les variables réalisées pas à pas ;
- En cas de résultats pas nettement tranchés dans l'analyse, de recourir à une analyse plus « *clinique* » des formes que prennent les distributions conjointes de données ;
- de manière générale, d'être sensible à la question de l'artéfact et des interprétations erronées possibles, liées surtout à la grande diversité, notamment en sciences humaines, des formes de relation entre variables, traduites par des formes variées de distributions conjointes de données ;
- de savoir, dans l'analyse les résultats, tenir compte des contextes et du sens que peuvent donner ces contextes aux résultats obtenus ;
- de pouvoir saisir, sous des formes complexes de distribution de données, ou derrière des tendances peu différenciées des indices obtenus dans CHIC, la complexité sous jacente notamment en termes différentiels.

<sup>18</sup> Voir, à ce sujet la réalisation pratique faite par Chamsine, D., 2012 (à paraître).

- Enfin de pouvoir renvoyer, in fine, à des approches plus individualisées pour une analyse plus fine de l'ensemble des données, notamment en confrontant les résultats plus collectifs à des situations individuelles précisément décrites et interprétées (études de cas).

## Références

- [1] Bruchon-Schweitzer, M., (2002). *Psychologie de la santé. Modèles, concepts et méthodes*. Paris, Dunod.
- [2] Chamsine, D. (2012, à paraître). Thèse de doctorat, Université de Caen.
- [3] Gras, R. (2005), Panorama du développement de l'A.S.I. à partir de situations fondatrices, Actes des *Troisièmes Rencontres Internationales –A.S.I. Analyse statistique Implicative- Palermo (Italy)*, 6-9 octobre, 9-33.
- [4] Gras, R., Couturier, R., (2010). Spécificités de l'Analyse Statistique Implicative (A.S.I.) par rapport à d'autres mesures de qualité de règles d'association, *Quaterni di Ricerca in Didattica (Mathematics)*, Supplemento n.1, n.20, 2010, G.R.I.M. (Department of Mathematics, University of Palermo, Italy) A.S.I. 5 Proceedings 5-7- November 2010, 175-200.
- [5] Lautrey, J., (1980). *Classe sociale, milieu familial, intelligence*,. Paris : PUF.
- [6] Longeot, F., (1978). *Les stades opératoires de Piaget et les facteurs de l'intelligence*, Paris, PUF de Grenoble.
- [7] Pellois, C. (2003). *Enseignement et formation le développement de la personne*, tome 2 "Le complexe" Collection « recherches et innovations », éditions l'Harmattan.
- [8] Pellois, C. (2004). *Enseignement et formation le développement de la personne*, tome 3 "Le sensible et le rationnel" Collection « recherches et innovations », éditions l'Harmattan.
- [9] Pellois, C. (2005). Complexité et développement : reconnaître l'émergence de qualités nouvelles dans le domaine des conduites adaptatives. Difficultés et intérêts des nouvelles perspectives. In : *La complexité ses formes, ses traitements, ses effets*, Coord. : Héryn, R., Cadet, B., *Actes du colloque 19-20 septembre 2002, Cahiers de la MRSH-Caen*, N° spécial, avril, 45-68.
- [10] Pellois, C. (2007). L'utilisation des modèles structuraux afin d'aborder la complexité du réel en sciences humaines, *Mathématiques et Sciences humaines*, 45<sup>ème</sup> année, n° 177(1), 53-85.
- [11] Pellois, C. (2008). Contrainte et liberté du sujet : entre incertitude et prévisibilité ? In *Cognition, incertitude et prévisibilité*, B. Cadet, G. Chasseigne, G. Foliot, Editions Publibook, Sciences Humaines et Sociales, Coll. Psychologie cognitive, Paris, 77-100.
- [12] Pellois, C. (2010). Sens et incertitude, une forme de complexité en psychologie : des contraintes aux parts de liberté, le développement et ses contextes, in Cadet, C., Chasseigne, G. *Traitement de la complexité dans les sciences humaines*, Editions Publibook Université, Coll. Psychologie Scientifique, 177-217.
- [13] Ribeaupierre, A., de, (1980). Application d'un modèle néo-piagétien à l'étude du stade des opérations formelles, *Bulletin de psychologie*, tome XXXIII, n° 345.