

# REVUE DE STATISTIQUE APPLIQUÉE

JEAN-CLAUDE RÉGNIER

RÉGIS GRAS

## **Statistique de rangs et analyse statistique implicite**

*Revue de statistique appliquée*, tome 53, n° 1 (2005), p. 5-38

[http://www.numdam.org/item?id=RSA\\_2005\\_\\_53\\_1\\_5\\_0](http://www.numdam.org/item?id=RSA_2005__53_1_5_0)

© Société française de statistique, 2005, tous droits réservés.

L'accès aux archives de la revue « *Revue de statistique appliquée* » (<http://www.sfds.asso.fr/publicat/rsa.htm>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

## STATISTIQUE DE RANGS ET ANALYSE STATISTIQUE IMPLICATIVE

Jean-Claude RÉGNIER (\*), Régis GRAS (\*\*)

(\*) EA3727 «savoirs, diversité, professionnalisation : e-learning, apprentissage via les supports numériques et les réseaux», Université Lumière Lyon II, Lyon France  
jean-claude.regnier@univ-lyon2.fr

(\*\*) École Polytechnique de l'Université de Nantes et LINA :  
Laboratoire Informatique de Nantes  
regisgra@club-internet.fr

### RÉSUMÉ

Le but de cet article est de discuter de l'apport d'une méthode d'analyse de données, à savoir l'analyse statistique implicative prise selon la conception de R. Gras (Gras 1979, 1996, 2000), à l'étude de la concordance-discordance des rangs accordés par des juges à des objets (ou variables qui peuvent être attributs, opinions, comportements, critères, etc.). Cette étude de concordance-discordance est à comprendre au sens de Friedman (Friedman 1937) ou de Kendall (Kendall, Babington Smith, 1939). Plus particulièrement, selon la même conception implicative, nous comparerons une analyse de préférences de ces juges exprimées par les rangs, avec l'extension de l'analyse implicative conçue par J. B. Lagrange (Lagrange, 1998), qu'il nomme propension entre variables à valeurs dans  $[0;1]$ , variables modales. Pour ce faire, nous nous affranchirons de l'hypothèse d'absence de lien *a priori* (indépendance) entre les variables. Mais nous pourrions cependant, d'une part, affecter d'une mesure de qualité des énoncés de la forme : «si l'objet *a* est rangé par les juges alors, généralement, l'objet *b* est rangé à un rang meilleur par les mêmes juges», d'autre part, représenter par un graphe les relations de préférences de l'ensemble des objets rangés. Dans cette étude, nous nous limitons aux deux cas des rangements complets et rangements incomplets mais sans *ex æquo* de *q* objets par *k* juges. Nous reportons à un prochain article, l'étude de la configuration qui prend en compte les *ex æquo*.

**Mots-clés :** Analyse implicative, propension, statistique de rangs, variables modales.

### ABSTRACT

This paper explores the benefit of one data analysis method implicative analysis such as R. Gras defined it to study ranks concordance-discordance accorded by judges to objects (or any variable such as attributes, opinions, behaviours, criteria, etc..) The meaning of concordance-discordance is the same as in Friedman's or Kendall's works. Then we compare one analysis of judges' preferences through ranks to the extension of implicative analysis built by Lagrange and which he calls propensity between modal variables. To achieve it, we don't consider the *a priori* independence hypothesis between variables. Nevertheless, we give a quality measure to "theorem" : "if an A object is ranked by judges, then generally a B object is ranked in a better way by the same judges" and we represent by graph relations of preferences for the

whole ranked objects. We limit this study to the two cases of complete ranking and incomplete ranking, without *ex æquo* between  $q$  objects by  $k$  judges. This special issue of *ex æquo* will be part of a further article.

**Keywords :** *Implicative analysis, propensity, Ranks statistics, modal variable.*

## Introduction

Pour introduire cet article, nous souhaitons ré-expliciter la problématique de la concordance des rangements de  $q$  objets par  $k$  ( $k > 2$ ) juges. Cela nous permet de poser nos notations au travers de la présentation du modèle que nous utilisons pour solutionner le problème soulevé par l'étude des accords-désaccords entre évaluateurs à propos des rangs attribués à une série d'objets.

Considérons  $q$  objets soumis au jugement individuel et indépendant de  $k$  individus d'une population  $E$  qui doivent fournir un rangement complet et sans *ex æquo* de ces objets. Il s'agit de savoir si on peut dégager de l'ensemble des  $k$  jugements, une concordance relativement à un rangement de ces  $q$  objets. C'est une des formes de l'analyse des préférences où, ici, nous n'envisageons pas de comparaison par paires.

La réalisation d'une telle opération peut être décrite par le tableau suivant :

TABLEAU 1

Objets	$O_1$	$O_2$	...	$O_j$	...	$O_q$
Individus						
1	$R_{11}^{(q)}$	$R_{12}^{(q)}$		$R_{1j}^{(q)}$		$R_{1q}^{(q)}$
2						
...						
i	$R_{i1}^{(q)}$	$R_{i2}^{(q)}$		$R_{ij}^{(q)}$		$R_{iq}^{(q)}$
...						
k	$R_{k1}^{(q)}$	$R_{k2}^{(q)}$		$R_{kj}^{(q)}$		$R_{kq}^{(q)}$

où chaque ligne  $i$  ( $i = 1, \dots, k$ ) est une permutation de  $(1, 2, \dots, q)$  en tant que réalisation du vecteur aléatoire  $(R_{ij}^{(q)})_{j=1, \dots, q}$ . Il y a donc  $q!$  réalisations possibles par ligne.

Les contraintes de rangement complet et sans *ex æquo* conduisent à ce que chaque ligne  $i$  correspondant à l'individu  $i$  vérifie la relation  $\sum_{j=1}^q R_{ij}^{(q)} = \frac{q(q+1)}{2}$

Sous l'hypothèse  $H_0$ , généralement légitime, de l'équiprobabilité des  $q!$  rangements des  $q$  objets par l'individu  $i$ , la probabilité pour l'objet  $O_j$ ,  $j \in \{0, 1, \dots, q\}$ ,

d'être placé au rang  $r$ ,  $r \in \{0, 1, \dots, q\}$  est de  $\frac{1}{q}$ . La probabilité pour que l'objet  $O_{j_1}$ ,  $j_1 \in \{0, 1, \dots, q\}$ , soit placé au rang  $r_1$ ,  $r_1 \in \{0, 1, \dots, q\}$ , et l'objet  $O_{j_2}$ ,  $j_2 \in \{0, 1, \dots, q\} - \{j_1\}$ , soit placé au rang  $r_2$ ,  $r_2 \in \{0, 1, \dots, q\}$  et  $r_2 \neq r_1$  est de  $\frac{1}{q(q-1)}$ .

Les caractéristiques des variables aléatoires  $R_{ij}^{(q)}$  sont :

TABLEAU 2

Espérance	$E(R_{ij}^{(q)}) = \frac{q+1}{2}$	Covariance	$Cov(R_{ij_1}^{(q)}; R_{ij_2}^{(q)}) = -\frac{q+1}{12}$
Variance	$V(R_{ij}^{(q)}) = \frac{q^2-1}{12}$	Corrélation	$Corr(R_{ij_1}^{(q)}; R_{ij_2}^{(q)}) = -\frac{1}{q-1}$

Autres propriétés :

TABLEAU 3

$E([R_{ij}^{(q)}]^2) = \frac{(q+1)(q+2)}{6}$	$E(R_{ij_1}^{(q)} R_{ij_2}^{(q)}) = \frac{(q+1)(3q+2)}{12}$
--	---

TABLEAU 4

$E(R_{ij_1}^{(q)} [R_{ij_2}^{(q)}]^2) = \frac{q(q+1)^2}{6}$
$E([R_{ij_1}^{(q)} R_{ij_2}^{(q)}]^2) = \frac{(q+1)(2q+1)(2q-1)(5q+6)}{180}$

### 1. Approche par les statistiques de Friedman et de Kendall

Reprenons le tableau 1 et considérons les sommes marginales dont les lignes sont constantes et égales à la somme des rangs attribués par un même individu.

Un tel tableau comporte  $(q!)^k$  réalisations équiprobables sous l'hypothèse  $H_0$  d'équiprobabilité des permutations caractérisant le rangement des  $q$  objets effectué indépendamment par chaque individu  $i$ .

Nous désignons par  $S(O_j)$  la somme des rangs attribués respectivement par les  $k$  individus à l'objet  $O_j$ . Pour connaître la distribution de probabilité exacte de  $S(O_j)$ , il suffirait de construire les  $(q!)^k$  tableaux. Une telle opération reste concrètement impossible dans les cas usuellement intéressants :  $k \geq 3$  et  $q \geq 5$  ce qui donnerait déjà plus de  $(5!)^3 = 1\,728\,000$  tableaux à construire.

TABLEAU 5

Objets Individus	$O_1$	$O_2$	...	$O_i$	...	$O_q$	Somme
<b>1</b>	$R_{11}^{(q)}$	$R_{12}^{(q)}$		$R_{1j}^{(q)}$		$R_{1q}^{(q)}$	$\frac{q(q+1)}{2}$
<b>2</b>							$\frac{q(q+1)}{2}$
...							
<b>i</b>	$R_{i1}^{(q)}$	$R_{i2}^{(q)}$		$R_{ij}^{(q)}$		$R_{iq}^{(q)}$	$\frac{q(q+1)}{2}$
...							
<b>k</b>	$R_{k1}^{(q)}$	$R_{k2}^{(q)}$		$R_{kj}^{(q)}$		$R_{kq}^{(q)}$	$\frac{q(q+1)}{2}$
<b>Somme</b>	$S(O_1)$	$S(O_2)$		$S(O_j)$		$S(O_q)$	$k \frac{q(q+1)}{2}$

Sous l'hypothèse  $H_0$ , nous pouvons cependant connaître les caractéristiques des variables  $S(O_j)$ ,  $j \in \{0, 1, \dots, q\}$ .

$$E(S(O_j)) = \sum_{i=1}^k E(R_{ij}^{(q)})$$

et

$$V(S(O_j)) = E \left[ \left( S(O_j) - k \frac{q+1}{2} \right)^2 \right] = E \left( \left[ \sum_{i=1}^k \left( R_{ij}^{(q)} - \frac{q+1}{2} \right) \right]^2 \right)$$

TABLEAU 6

Espérance	$E(S(O_j)) = k \frac{(q+1)}{2}$
Variance	$V(S(O_j)) = k \frac{q^2 - 1}{12}$

### 1.1 Interprétation des $S(O_j)$ pour expliciter une concordance des rangements

Nous pouvons utiliser l'ordre des sommes des rangs attribués à chaque objet pour déterminer un rangement que nous interprétons comme étant celui qui «s'approche au mieux» du point de vue des juges.

Deux cas extrêmes peuvent apparaître.

*Concordance « parfaite »*

Cette situation correspond au cas où les juges attribueraient respectivement à chaque objet le même rang. Il y a  $q!$  tableaux de ce type sur les  $(q!)^k$  possibles. Sous  $H_0$ , cet événement a donc une probabilité  $\left(\frac{1}{q!}\right)^{k-1}$  de survenir.

À une permutation près, le vecteur  $(S(O_j))_{j=1,\dots,q}$  prend alors la valeur  $(k, 2k, 3k, \dots, qk)$ . Dans une étude statistique conduisant à cette situation de concordance parfaite, il est clair que nous pouvons rejeter l'hypothèse  $H_0$  d'indépendance des jugements entre les juges au profit de l'hypothèse alternative de l'existence d'une tendance à l'accord pour un rangement des objets déterminé par cette permutation, ici  $(O_1, O_2, O_3, \dots, O_q)$ . En effet, un tel événement admet une probabilité de réalisation de  $\left(\frac{1}{q!}\right)^{k-1}$ . Cette  $p$ -valeur est évidemment très faible et, en tout cas, très inférieure au seuil  $\alpha$  de risque de 1<sup>ère</sup> espèce de l'ordre de 0,05 ou 0,01, habituellement posé par le chercheur

*Hétérogénéité « parfaite » des rangements*

Cette situation correspond au cas où l'attribution des rangs aux  $q$  objets produirait par des effets de compensation des sommes de rangs égales ou presque égales à une unité près.

En effet nous rappelons que la somme des  $S(O_j)_{j=1,\dots,q}$  est constante et égale à  $k \frac{q(q+1)}{2}$ .

Ainsi en cas d'égalité de ces  $q$  sommes, leur valeur commune devrait être  $k \frac{(q+1)}{2}$ . Si le produit  $k(q+1)$  est pair, alors il existe effectivement des tableaux dont le vecteur  $(S(O_j))_{j=1,\dots,q}$  prend la valeur  $(k \frac{(q+1)}{2}, k \frac{(q+1)}{2}, k \frac{(q+1)}{2}, \dots, k \frac{(q+1)}{2})$ . Si le produit  $k(q+1)$  est impair – ce qui ne peut se produire que si  $k$  est impair et  $q$  est pair – alors  $\frac{q}{2}$  composantes du vecteur  $(S(O_j))_{j=1,\dots,q}$  prendront la valeur  $k \frac{(q+1)}{2} - \frac{1}{2}$  tandis que les  $\frac{q}{2}$  autres composantes vaudront  $k \frac{(q+1)}{2} + \frac{1}{2}$ .

Le nombre de tableaux vérifiant ces conditions est à l'évidence très supérieur aux  $q!$  tableaux de la concordance « parfaite ». Intuitivement, sous  $H_0$ , nous avons beaucoup plus de chance d'obtenir un tableau du type « hétérogénéité parfaite » que du type « concordance parfaite ».

Si nous traduisons ce résultat sous la forme d'un rangement basé sur l'ordre des sommes  $(S(O_j))_{j=1,\dots,q}$ , nous obtenons un rangement des  $q$  objets constitué par  $q$  rangs *ex æquo*<sup>1</sup> :  $(\frac{(q+1)}{2}, \frac{(q+1)}{2}, \frac{(q+1)}{2}, \dots, \frac{(q+1)}{2})$  ou encore à une permutation près en deux groupes de  $\frac{q}{2}$  objets rangés ainsi :  $\frac{q}{2}$  objets *ex æquo* occupent

<sup>1</sup> En ce concerne le traitement des *ex æquo*, nous utilisons la règle suivante qui respecte la contrainte de la somme des  $q$  rangs égale à  $\frac{q(q+1)}{2}$  : les rangs attribués aux *ex æquo* sont la moyenne des rangs occupés par les objets s'il n'y avait pas eu d'*ex æquo*.

les rangs 1 à  $\frac{q}{2}$  et les  $\frac{q}{2}$  autres objets, les rangs  $\frac{q}{2} + 1$  à  $\frac{q}{2} + \frac{q}{2}$ , d'où le vecteur de rangs suivant  $\left(\frac{\frac{q}{2} + 1}{2}, \dots, \frac{\frac{q}{2} + 1}{2}, \frac{\frac{3q}{2} + 1}{2}, \dots, \frac{\frac{3q}{2} + 1}{2}\right)$

Entre «concordance parfaite» et «hétérogénéité parfaite»

La question demeure quant à l'interprétation des cas intermédiaires. Quelle décision prendre quand nous obtenons des résultats compris entre ces deux extrêmes? La statistique  $F_{q,k}$  de Friedman ou celle  $W$  de Kendall constituent des outils d'aide à la prise de cette décision.

$$F_{q,k} = \frac{12k}{q(q+1)} \sum_{j=1}^q \left( \frac{S(O_j)}{k} - \frac{q+1}{2} \right)^2$$

et

$$W = \frac{12 \sum_{j=1}^q \left( S(O_j) - k \frac{(q+1)}{2} \right)^2}{k^2(q^3 - q)}$$

Ces deux statistiques sont reliées par :  $F_{q,k} = k(q-1)W$ . Elles constituent des mesures d'écart entre la situation observée et celle de l'hétérogénéité «parfaite».

Si nous posons  $S_K = \sum_{j=1}^q \left( S(O_j) - k \frac{(q+1)}{2} \right)^2$ , nous constatons que  $S_K$  est minimum dans le cas de l'hétérogénéité «parfaite» -  $S_{K \min} = 0$  ou  $S_{K \min} = \frac{q}{4}$  - et  $S_K$  est maximum dans le cas de la concordance «parfaite» -  $S_{K \max} = \frac{k^2(q^3 - q)}{12}$  - Ainsi la statistique  $W = \frac{S_K}{S_{K \max}}$  donne un indice compris entre 0 et 1 qui traduit un degré de concordance entre les  $k$  juges relativement au rangement des  $q$  objets.

Un autre approche pourrait être réalisée à partir de la confrontation des  $C_k^2$  paires  $\{i_1, i_2\}$  de juges. Les  $\frac{k(k-1)}{2}$  coefficients empiriques de Spearman  $R_S(i_1, i_2) =$

$1 - \frac{6 \sum_{j=1}^q \left( R_{i_1 j}^{(q)} - R_{i_2 j}^{(q)} \right)^2}{q(q^2 - 1)}$  fourniraient une information dont l'interprétation globale

resterait toutefois difficile à réaliser. La moyenne  $\overline{R_S}$  de ces  $\frac{k(k-1)}{2}$  coefficients empiriques de Spearman donnerait une information plus synthétique. En fait il existe un lien algébrique entre  $\overline{R_S}$  et  $W$  exprimé par la relation :  $W = \frac{(k-1)\overline{R_S} + 1}{k}$  (Lecoutre & Tassi, 1987, p. 226).

Notons que la procédure adoptée qui consiste à ranger les  $q$  objets selon l'ordre déterminé par les  $q$  sommes  $S(O_j)_{j=1 \text{ à } q}$ , possède la propriété de maximiser la

somme des  $k$  valeurs des coefficients de Spearman entre le rangement synthétique obtenu et les  $k$  rangements individuels. (Saporta 1990 p. 147)

### 1.2. Test de concordance des rangements de $q$ objets par $k$ juges, $W$ de Kendall

La loi de probabilité de  $W$  sous  $H_0$  ne dépend que des  $(q!)^k$  résultats possibles. Elle est théoriquement connaissable de façon exacte mais à l'évidence pratiquement impossible en raison de la taille des calculs en jeu.

Sous l'hypothèse  $H_0$ , nous pouvons obtenir les caractéristiques suivantes (Lecoutre & Tassi 1987 p. 229-230) :

TABLEAU 7  
Moments de  $W$  sous  $H_0$

Espérance	$E_{H_0}(W) = \frac{1}{k}$
Variance	$V_{H_0}(W) = \frac{2(k-1)}{k^3(q-1)}$
Moment centré d'ordre 3	$\mu_{3H_0}(W) = \frac{8(k-1)(k-2)}{k^5(q-1)^2}$
Moment centré d'ordre 4	$\mu_{4H_0}(W) = \frac{12(k-1)^2}{k^6(q-1)^2} + \frac{48(k-1)(k-2)(k-3)}{k^7(q-1)^2} - \frac{48(k-1)}{k^7(q+1)(q-1)^3}$

Le test  $W$  de Kendall revient à confronter l'hypothèse  $H_0$  d'hétérogénéité des  $k$  jugements – d'absence de point de vue concordant – à l'égard d'un rangement des  $q$  objets, à l'hypothèse  $H_1$  d'homogénéité des  $k$  jugements – d'existence d'une tendance à la concordance des juges – relativement au rangement de ces  $q$  objets. Il s'effectue à partir de la confrontation de la réalisation de  $\bar{W}$  issue de l'observation, à une valeur critique  $c_W(\alpha)$  telle que  $\text{Prob}_{H_0}\{W > c_W(\alpha)\} \leq \alpha$ , où  $\alpha$  est le niveau du risque de 1<sup>ère</sup> espèce. Cette valeur critique est obtenue à partir d'une table lorsque  $q < 7$  (Saporta 1990 p. 464) (Lecoutre & Tassi, 1987, p. 427-431). Pour  $3 \leq q \leq 7$  et  $3 \leq k \leq 20$ , on peut utiliser la statistique  $\frac{(k-1)W}{1-W}$  dont la loi est approchée



par celle de la variable de Fischer-Snedecor  $F\left(n-1-\frac{2}{k}; (k-1)n\left(n-1-\frac{2}{k}\right)\right)$ . Sinon pour  $q \geq 7$ , on utilise la loi de probabilité de la statistique  $k(q-1)W$  qui est approximativement la loi de probabilité du  $\chi^2_{(q-1)}$  à  $q-1$  ddl. (Lecoutre & Tassi, 1987, p. 232)

Le rejet de  $H_0$  revient à considérer que la valeur, réalisation empirique de  $W$ , s'éloigne significativement de la valeur minimale correspondant à la situation d'hétérogénéité « parfaite » et s'approche de la valeur maximale 1 correspondant à une tendance à une concordance relativement au rangement déterminé par l'ordre des sommes de rangs attribués à chaque objet par les  $k$  juges.

Le non-rejet de  $H_0$  revient à ne pas renoncer à la situation d'hétérogénéité « parfaite » et à considérer qu'il ne se dégage pas de tendance à la concordance pour un rangement des  $q$  objets au sein du groupe des  $k$  juges. Évidemment cela expose au risque de 2<sup>nde</sup> espèce d'un niveau  $\beta$  inconnu puisque l'hypothèse alternative  $H_1$  est composite.

## 2. Analyse statistique implicative comme outil d'exploration de la structure des rangements (cas du rangement complet et sans *ex aequo*)

À l'origine, l'analyse statistique implicative s'applique aux variables binaires en produisant des règles, puis aux classes de variables binaires en produisant des méta-règles. Depuis, d'autres types de variables, comme des variables-intervalles par exemple, ont été considérées. Cette méthode d'analyse de données constitue un outil pour expliciter statistiquement des structures quasi-implicatives – permettant d'exprimer le fait que « des individus qui possèdent un caractère  $a$  ont tendance à posséder aussi le caractère  $b$  ». Ainsi peut-on construire un modèle statistique permettant de donner un cadre à des propriétés par exemple du type : « Réussir à résoudre des problèmes arithmétiques multiplicatifs en cycle 3 » implique généralement « Réussir à résoudre des problèmes arithmétiques additifs en cycle 3 », sans pour autant qu'un sens causal n'accompagne nécessairement cette « tendance implicative ». Le tableau 8 montre la répartition des occurrences dans chaque cas.  $N(\dots)$  représentant la variable CARDINAL,  $N(a)$  est alors une réalisation empirique qui donne le nombre d'individus de l'ensemble  $E$ , population-mère, de taille  $N(E) = N$ , qui possèdent le caractère  $a$ , c'est-à-dire le cardinal du sous-ensemble  $A$  de  $E$  des individus qui possèdent le caractère  $a$ . Dans ces conditions,  $N(a \wedge \neg b)$  est le nombre d'individus qui possèdent le caractère  $a$  mais pas le caractère  $b$ .

TABLEAU 8

		Caractère b		
		1	0	
Caractère a	1	$N(a \wedge b)$	$N(a \wedge \neg b)$	$N(a)$
	0	$N(\neg a \wedge b)$	$N(\neg a \wedge \neg b)$	$N(\neg a)$
		$N(b)$	$N(\neg b)$	$N$

Nous supposons que  $N(a) < N(b)$  et nous nous intéressons aux individus qui contredisent l'implication mathématique dont l'effectif est  $N(a \wedge \neg b)$ . Ce choix s'appuie en particulier sur le fait qu'il est plus aisé de considérer ce cas que les trois autres cas pour évaluer l'implication. Sous l'hypothèse  $H_0$  d'absence de lien *a priori* entre les deux caractères  $a$  et  $b$ , l'effectif théorique espéré serait de  $\frac{N(a)N(\neg b)}{N}$ . Nous savons que  $0 \leq (N(a \wedge \neg b) \leq \min\{N(a); N(\neg b)\}$ , mais nous pouvons aussi interpréter :

$0 \leq N(a \wedge \neg b) < \frac{N(a)N(\neg b)}{N}$	L'effectif observé étant inférieur à l'effectif théorique signifie que la dépendance entre la présence du caractère ( $a$ ) et l'absence du caractère ( $b$ ) est répulsive. En terme d'implication statistique, nous interprétons une observation qui va dans le sens de l'acceptation de la règle ( $a \implies b$ )
$0 < \frac{N(a)N(\neg b)}{N} < N(a \wedge \neg b)$	L'effectif observé étant supérieur à l'effectif théorique signifie que la dépendance entre la présence du caractère ( $a$ ) et l'absence du caractère ( $b$ ) est attractive. En terme d'implication statistique, nous interprétons une observation qui va dans le sens du rejet de la règle ( $a \implies b$ )

À la base de la théorie, R. Gras (Gras 1996 p. 32) définit alors l'*indice d'implication statistique* par centrage et réduction de la variable  $N(\dots)$  donnant le nombre de contre-exemples à l'implication statistique  $a \implies b$  et dont l'espérance vaut  $E[N] = \frac{N(a)N(\neg b)}{N}$  sous l'hypothèse d'indépendance des deux caractères ( $a$ ) et ( $b$ ) :

$$q(a, \neg b) = \frac{N(a \wedge \neg b) - \frac{N(a)N(\neg b)}{N}}{\sqrt{\text{var}[N(\dots)]}}$$

réalisation de la statistique  $Q(a, \neg b)$  dont la loi de probabilité est établie à partir de celle de la variable  $N(\dots)$ . Nous pouvons constater que la variable  $N(\dots)$  a été explicitée selon trois modèles faisant respectivement référence à une loi hypergéométrique, une loi binomiale (Lerman, Gras, Rostam 1981), ou même à une loi de Poisson paramètre  $\lambda = \frac{N(a)N(\neg b)}{N}$  (Gras 1996 p. 30-31). Sous les conditions de taille suffisante de l'ensemble de référence, la loi de la statistique  $Q(a, \neg b)$  est approchée par celle de la variable de Laplace-Gauss  $LG(0; 1)$  centrée réduite (Gras 1996 p. 32). Nous retiendrons ici le modèle binomial comme R. Gras l'avait d'ailleurs initialement fait dans sa thèse (Gras, 1979 p. 51). Pour ce faire

nous considérons les deux sous-ensembles  $A$  et  $B$  de l'ensemble  $E$  tels que  $\text{Card}A = N(a)$  et  $\text{Card}B = N(b)$ , qui correspondent à l'observation du phénomène étudié. Nous considérons ensuite deux sous-ensembles  $X$  et  $Y$  obtenus indépendamment et aléatoirement par tirage avec remise dans  $E$  tels que  $\text{Card}X = N(a)$  et  $\text{Card}Y = N(b)$ . Dans ces conditions la probabilité d'appartenance à  $X$  mais pas à  $B$  est égale la probabilité d'appartenance à  $A$  mais pas à  $Y$ , c'est-à-dire  $\pi = \frac{N(a)N(-b)}{N^2}$ . De là, la loi de probabilité

$$\text{Prob}\{\text{Card}(A \cap \neg Y) = k\} = \text{Prob}\{\text{Card}(X \cap \neg B) = k\} = \binom{N}{k} \pi^k (1 - \pi)^{N-k}$$

$N(a \wedge \neg b)$  représente une réalisation empirique de l'une ou l'autre de ces deux variables. Dans cette perspective : *L'implication statistique  $a \implies b$  est admissible à un niveau de confiance  $1 - \alpha$ , si et seulement si  $\text{Prob}\{\text{Card}(A \cap \neg Y) \leq N(a \wedge \neg b)\} \leq \alpha$ .*

*L'intensité d'implication statistique  $\varphi(a, \neg b)$  est alors définie à partir de la probabilité que le nombre de contre-exemples observé soit inférieur à celui que donnerait le hasard, c'est-à-dire en posant  $Q(a, \neg b) = \frac{\text{Card}(A \cap \neg Y) - N\pi}{\sqrt{N\pi(1 - \pi)}}$  =*

$$\frac{\text{Card}(A \cap \neg Y) - \frac{N(a)N(-b)}{N}}{\sqrt{\frac{N(a)N(-b)}{N} \left(1 - \frac{N(a)N(-b)}{N^2}\right)}}$$

$$\varphi(a, \neg b) = 1 - \text{Prob}[Q(a, \neg b) \leq q(a, \neg b)] = 1 - \sum_{k=0}^{k=N(a, \neg b)} \binom{N}{k} \pi^k (1 - \pi)^{N-k}$$

Ainsi *l'intensité d'implication statistique  $\varphi(a, \neg b)$  conduit à formuler la définition suivante : L'implication statistique  $a \implies b$  est admissible à un niveau de confiance  $1 - \alpha$ , si et seulement si  $\varphi(a, \neg b) \geq 1 - \alpha$ .*

Le rapprochement entre variables de rangs et analyse statistique implicative apparaît avec l'extension de cette théorie aux variables numériques et classes de variables numériques prenant des valeurs dans l'intervalle  $[0; 1]$  de  $\mathbb{R}$ . De telles variables sont nommées : variables modales. Pour illustrer cette modélisation, prenons l'exemple suivant d'un énoncé soumis au jugement d'un étudiant :

«Quand j'entre en cours de statistique, je suis complètement paniqué(e)»

Cet énoncé correspond-il à ce que vous éprouvez? non=0; oui=1

Modèle «variable binaire» : non=0; oui=1

Modèle «variable modale» : pas du tout= 0; un peu= 1/3; oui en partie= 2/3; tout à fait= 1

Dans ce passage des variables binaires aux variables modales, J.B. Lagrange (Lagrange 1998) introduit la notion de «propension entre deux variables à valeurs dans l'intervalle  $[0; 1]$ ». Il définit alors un *indice (coefficient) aléatoire de propension*

$\tilde{Q}(a, -b)$  d'une variable modale  $a$  vers une variable modale  $b$  ainsi qu'une *intensité de propension*  $\Phi(-\tilde{q}(a, -b))$  où  $\tilde{q}(a, -b)$  est une réalisation de la statistique  $\tilde{Q}(a, -b)$  et  $\Phi$  est la fonction de répartition de la variable de Laplace-Gauss centrée réduite LG(0, 1). Le modèle qu'il adopte – que nous rappellerons plus loin – est une stricte extension de l'implication statistique dans la mesure où J.B. Lagrange montre que si  $a$  et  $b$  sont des variables binaires alors  $\tilde{q}(a, -b) = q(a, -b)$ .

Dans le cadre de l'approche par l'analyse statistique implicative, le recours aux variables de rangs apparaît dans les travaux de M. Bailleul (Bailleul 1994). Cherchant à explorer les représentations de l'enseignement des mathématiques, il propose le protocole suivant, inspiré de la technique du  $Q$ -sort, pour construire certaines de ses données : les enseignants de son échantillon doivent choisir, parmi une liste de 30 mots, les trois qui caractérisent le mieux l'enseignement des mathématiques selon un certain point de vue. Ensuite ces trois mots doivent être rangés selon les rangs 1, 2, 3 d'importance décroissante. La réponse est un des  $A_{30}^3 = \frac{30!}{27!} = 24\,360$  arrangements (tiercés) possibles. Dans son choix de l'outil mathématique d'analyse, M. Bailleul inscrit la contrainte de la prise en compte d'un indice non-symétrique lui permettant de répondre à la question : *dans quelle mesure peut-on affirmer, de façon relativement sûre, que le choix par un individu de telle variable va entraîner le choix de telle autre variable?* Et concernant le cas des variables de rangs qui nous intéresse : *dans quelle mesure le choix de telle variable avec tel rang entraîne-t-il le choix de telle autre variable avec un rang meilleur ou égal à celui affecté à la première?*

### 2.1. Transformation d'une variable de rangs en une variable modale

Nous nous replaçons dans le cas d'un rangement complet et sans *ex æquo* tel que nous l'avons abordé en première partie de cet article.

En général il est demandé aux  $k$  individus de ranger les  $q$  objets par ordre de préférence décroissante. Ainsi le rang 1 est le plus important tandis que le rang  $q$  correspond à l'objet le moins important. Or dans le modèle proposé par J.B Lagrange la valeur 1 représente le plus haut degré d'adhésion possible tandis que le degré minimum d'adhésion est coté par 0. Dans ces conditions, la transformation suivante est

une solution à ce problème :  $\mathbf{X}_{ij}^{(q)} = \frac{q - \mathbf{R}_{ij}^{(q)}}{q - 1}$  qui respecte la contrainte sémantique. En effet cette transformation de  $\{1, 2, \dots, q\}$  vers  $[0; 1]$  donne la correspondance suivante (Tableau 9)

Il est évident que si le rangement était effectué dans le sens de préférence croissante, c'est alors la transformation  $\mathbf{X}'_{ij}^{(q)} = \frac{\mathbf{R}_{ij}^{(q)} - 1}{q - 1}$  qu'il faudrait appliquer.

Dans les deux cas, cela revient alors à demander aux  $k$  individus de réaliser une des  $q!$  bijections entre les  $q$  objets et l'ensemble suivant des degrés d'adhésion  $\{0, \frac{1}{q-1}; \frac{2}{q-1}; \dots; \frac{q-2}{q-1}; 1\}$ . Cette opération risquerait toutefois de générer des biais dus à une surcharge mnésique du sujet interrogé. Le recours aux nombres ordinaux reste un moyen plus proche des pratiques quotidiennes de classements ordonnés.

Dans la suite, nous poserons  $m_j$  la moyenne empirique des valeurs reçues par l'objet  $O_j$ . La relation entre cette moyenne et la moyenne empirique  $\frac{S(O_j)}{k}$  de la variable rang pour un objet  $O_j$  est alors :

$$m_j = \frac{1}{k} \sum_{i=1}^{i=k} X_{ij}^{(q)} = \frac{q}{q-1} - \frac{S(O_j)}{k(q-1)}$$

TABLEAU 9

Rang attribué par l'individu à l'objet $O_j$ parmi les $q$ objets	Degré d'adhésion exprimé par le rang choisi sur $[0; 1]$
1	1
2	$\frac{q-2}{q-1}$
...	...
$q-1$	$\frac{1}{q-1}$
$q$	0

## 2.2. Le modèle de la quantification de la propension de J.B. Lagrange

Étant données deux variables modales  $A$  et  $B$ ,  $(a_i)_{i=1,\dots,k}$  et  $(b_i)_{i=1,\dots,k}$  deux vecteurs de  $[0; 1]^k$  représentant une réalisation de ces variables auprès des  $k$  individus d'une population  $P$ . «Il y a propension de  $A$  vers  $B$  si on rencontre en moyenne peu d'individus pour lesquels  $(a_i)$  est fort et  $(b_i)$  l'est moins. La moyenne sur  $P$  des  $a_i(1-b_i)$  est prise comme un indice de non-propension de  $A$  vers  $B$ . Cette valeur moyenne est comparée à celle qui résulterait d'une hypothèse d'absence de lien entre  $A$  et  $B$ .» J.B. Lagrange propose alors la statistique «coefficient de propension» suivante (Lagrange 1998) :

- soit, d'une part,  $E$  un ensemble de «grande» taille  $n$ , et un ensemble de variables aléatoires  $(A_i)_{i \in E}$  et  $(B_i)_{i \in E}$  de même loi, représentatives des variables  $A$  et  $B$ ;
- soit, d'autre part, une partie  $P$  de taille  $k$  aléatoirement extraite de  $E$ . Les  $k$  événements  $\{i \in P\}$  sont indépendants et équiprobables de probabilité  $\frac{k}{n}$ .

La variable aléatoire  $Z = \frac{\sum_{i \in P} A_i(1-B_i)}{k}$  admet comme réalisation l'indice moyen de non-propension  $s = \frac{\sum_{i \in P} a_i(1-b_i)}{k}$ .

En posant  $T_i = 1_{\{i \in P\}} A_i (1 - B_i)$ ,  $Z$  peut encore être exprimée par  $Z = \frac{\sum_{i \in E} T_i}{k}$ . En appliquant le théorème central limite, J.B. Lagrange démontre que la variable  $\tilde{Q}(a, -b) = \frac{Z - E[A_i(1 - B_i)]}{\sqrt{\frac{E([A_i(1 - B_i)]^2)}{k}}}$  est approximativement distribuée comme

la variable LG(0; 1) de Laplace Gauss centrée réduite. En prenant aussi en compte la propriété d'indépendance des variables  $A_i$  et  $B_i$  pour tout  $i$ , JB Lagrange parvient à la définition suivante :

Soit  $A$  et  $B$  deux variables modales respectivement de moyenne empirique  $m_a$  et  $m_b$ , et de variance empirique  $\nu_a$  et  $\nu_b$ , on appelle *coefficient de propension de A vers B* le nombre :

$$\tilde{q}(a, -b) = \frac{\frac{\sum_{i \in P} a_i(1 - b_i)}{k} - m_a(1 - m_b)}{\sqrt{\frac{(\nu_a + m_a^2)(\nu_b + (1 - m_b)^2)}{k}}}$$

et *intensité de propension de A vers B*, le nombre  $\Phi(-\tilde{q}(a, -b)) = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\tilde{q}(a, -b)} e^{-\frac{t^2}{2}} dt$

En fait, J.B. Lagrange utilise les valeurs empiriques de la moyenne et de la variance pour estimer l'espérance et la variance de la statistique  $Z$ .

Par ailleurs il introduit la quantité  $\frac{m_x(1 - m_x)}{\nu_x}$  nommée « facteur de variance » en tant qu'information qui permet de fournir une condition suffisante relative à la compatibilité entre la relation de propension et la relation d'ordre sur les moyennes.

### 2.3. Application du modèle de la propension aux variables de rangs

Revenons au tableau 1. L'exploration de la structure de ce tableau par l'analyse statistique de la propension requiert un réexamen des conditions d'application du modèle de J.B. Lagrange. Afin de pouvoir comparer les approches, nous supposons que les valeurs modales sont réparties de façon régulière (équidistante) sur  $[0; 1]$ . Les  $k$  événements  $\{i \in P\}$  qui définissent les lignes du tableau, sont évidemment supposés *a priori* indépendants. Là où un problème nous paraît surgir, dans le cadre du traitement modal en terme de rangs, c'est dans le choix *a priori* d'une hypothèse d'indépendance entre les deux variables  $A_i$  et  $B_i$ . Dans la procédure de rangement comme nous l'avons précisé plus haut, chaque composante du vecteur  $i$ , c'est-à-dire chaque variable aléatoire rattachée au critère  $i$  et à l'objet  $j$ , dépend des  $q - 1$  autres objets. Il n'est donc pas possible ici de considérer l'hypothèse d'indépendance de deux variables  $A_i$  et  $B_i$  – dont la réalisation donne respectivement le rang de l'objet  $A$  et celui de l'objet  $B$  – comme acceptable. Elle est en effet incompatible avec la procédure même qui conduit l'individu à se confronter à la totalité des  $q$  objets pour

produire finalement une permutation de  $\{1, \dots, q\}$ . Ceci nous montre qu'il n'est pas possible de transporter le modèle de Lagrange dans notre modèle de rangements complets sans *ex æquo*.

Comment peut-on exprimer la structure issue du rangement complet et sans *ex æquo* des  $q$  objets par  $k$  sujets en terme de relation de propension?

Essayons de réinterpréter la quantité  $s = s(A \implies B) = \frac{\sum_{i \in P} a_i(1-b_i)}{k}$  dans le contexte des variables de rangs. Dire que  $a_i$  est fort, traduit le fait que l'objet  $A$  considéré est placé à un haut degré de préférence. Dire que  $b_i$  est faible, traduit le fait que l'objet  $B$  considéré est placé à un bas degré de préférence. La relation de propension de  $a$  vers  $b$  devrait traduire le fait que «la tendance à placer l'objet  $A$  à un rang  $r_A$  entraîne celle à placer l'objet  $B$  à un rang  $r_B$  correspondant à une préférence de  $B$  sur  $A$ .» Dans notre modélisation du choix préférentiel,  $B$  est préféré à  $A$  si  $\text{rang}(B) < \text{rang}(A)$  et donc si  $0 \leq \frac{q - \text{rang}(A)}{q - 1} < \frac{q - \text{rang}(B)}{q - 1} \leq 1$ .

Dans ces conditions, en conservant la statistique  $\tilde{Q}(a, -b) = \frac{Z - E[A_i(1-B_i)]}{\sqrt{\frac{E([A_i(1-B_i)]^2)}{k}}}$ ,

il nous faut prendre en compte le fait que les variables  $A_i$  et  $B_i$  sont dépendantes pour le calcul de l'espérance et de la variance de  $Z$ . Nous conservons l'approximation gaussienne de la variable indice aléatoire  $Z$  dans la mesure où nous restons dans les conditions posées par Lagrange : d'une part, pour chaque  $i$ , l'évènement  $\{i \in P\}$  et les évènements définis par la variable  $A_i(1 - B_i)$  sont indépendants, d'autre part, les variables  $T_i$  sont indépendantes et identiquement distribuées.

#### 2.4. Espérance et variance de $Z$

Les calculs bien que sans difficulté majeure, demeurent fastidieux et demandent une certaine attention. Pour donner des repères dans ce calcul, nous fournissons quelques résultats intermédiaires

Calcul de  $E(Z) = E[A_i(1 - B_i)]$

Nous avons obtenu le résultat suivant

$$\mathbf{E}(A_i(1 - B_i)) = \frac{1}{2} - \frac{(q-2)(3q-1)}{12(q-1)^2}$$

En posant  $Y = A_i(1 - B_i)$ , nous obtenons  $E(Y) = E(A_i) - E(A_i B_i)$ .

Or  $A_i = \frac{q - R_{ij_1}^{(q)}}{q - 1}$  et  $B_i = \frac{q - R_{ij_2}^{(q)}}{q - 1}$  en considérant que  $A_i$  (resp.  $B_i$ ) représente la variable aléatoire d'attribution du degré d'importance par l'individu  $i$  à l'objet  $j_1$  (resp. à l'objet  $j_2$ ).

$$E(A_i) = \frac{q - E[\mathbf{R}_{i,j_1}^{(q)}]}{q-1} = \frac{q - \frac{q+1}{2}}{q-1} = \frac{1}{2} = E(B_i) \text{ et } E(A_i B_i) = \frac{(q-2)(3q-1)}{12(q-1)^2} \text{ nous permettent alors de d\u00e9duire l'esp\u00e9rance de } Z.$$

Calcul de  $E[(A_i(1 - B_i))^2]$

En partant de  $Y^2 = [A_i(1 - B_i)]^2$ , nous avons  $E(Y^2) = E[(A_i)^2] + E[(A_i B_i)^2] - E[2(A_i^2 B_i)]$ . En appliquant les r\u00e9sultats relatifs aux rangs nous obtenons :  $E[(A_i)^2] = \frac{2q-1}{6(q-1)}$ , puis  $E[(A_i B_i)^2] = \frac{(2q-1)(q-2)(2q-3)(5q+1)}{180(q-1)^4}$  et enfin  $E[(A_i^2 B_i)] = \frac{q(q-2)}{6(q-1)^2}$ .

De l\u00e0 nous en d\u00e9duisons que :

$$E[[A_i(1 - B_i)]^2] = \frac{20q^4 - 46q^3 + 49q^2 - 41q + 24}{180(q-1)^4}$$

### 2.5. Expression du coefficient de propension pour des variables de rangs

Dans ce contexte particulier, la valeur empirique du coefficient de propension s'exprime alors de la fa\u00e7on suivante :

$$\begin{aligned} \tilde{q}_{(q)}(a, -b) &= \frac{\sum_{i \in P} a_i(1 - b_i)}{k} - \left[ \frac{1}{2} - \frac{(q-2)(3q-1)}{12(q-1)^2} \right] \\ &= \frac{\sum_{i \in P} a_i(1 - b_i)}{k} - \left[ \frac{1}{4} + \frac{(q+1)}{12(q-1)^2} \right] \\ &= \frac{\sum_{i \in P} a_i(1 - b_i)}{k} - \frac{1}{4} \end{aligned}$$

On peut noter que si le nombre d'objets  $q$  tend vers l'infini, la valeur empirique du coefficient de propension tend vers  $\tilde{q}_{(\infty)}(a, -b) = \frac{\sum_{i \in P} a_i(1 - b_i)}{k} - \frac{1}{4}$ .

$$\frac{1}{3\sqrt{k}}$$



**2.6. Compatibilité de la relation de propension (permutationnelle)  
avec l'ordre des moyennes**

Considérons deux variables  $A$  et  $B$  issues de la transformation des rangs attribués aux objets  $O_1$  à  $O_q$ . Notons  $m_a$  et  $m_b$  leurs moyennes empiriques respectives obtenues sur l'échantillon des  $k$  juges et  $\Phi$  la fonction de répartition de  $LG(0; 1)$ . Pour mémoire, nous rappelons la relation entre la moyenne empirique  $\frac{S(O_j)}{k}$  de la variable modale et la moyenne empirique de la variable rang pour un objet  $O_j$  :

$$m_j = \frac{1}{k} \sum_{i=1}^{i=k} X_{ij}^{(q)} = \frac{q}{q-1} - \frac{S(O_j)}{k(q-1)}$$

*Proposition 1* :  $\Phi(-\tilde{q}_{(q)}(a, -b)) \geq \Phi(-\tilde{q}_{(q)}(b, -a))$  si et seulement si  $m_a \leq m_b$

En effet  $\Phi(-\tilde{q}_{(q)}(a, -b)) \geq \Phi(-\tilde{q}_{(q)}(b, -a)) \iff \tilde{q}_{(q)}(a, -b) \leq \tilde{q}_{(q)}(b, -a)$

De là, nous pouvons en déduire que :

$$\frac{\sum_{i \in P} a_i(1-b_i)}{k} - \frac{1}{2} + \frac{(q-2)(3q-1)}{12(q-1)^2} \leq \frac{\sum_{i \in P} b_i(1-a_i)}{k} - \left| \frac{1}{2} - \frac{(q+1)}{12(q-1)^2} \right|$$

$$\sqrt{\frac{20q^4 - 46q^3 + 49q^2 - 41q + 24}{180k(q-1)^4}} \leq \sqrt{\frac{20q^4 - 46q^3 + 49q^2 - 41q + 24}{180k(q-1)^4}}$$

et que  $\frac{\sum_{i \in P} a_i(1-b_i)}{k} \leq \frac{\sum_{i \in P} b_i(1-a_i)}{k} \iff \frac{\sum_{i \in P} a_i - \sum_{i \in P} a_i b_i}{k} \leq \frac{\sum_{i \in P} b_i - \sum_{i \in P} a_i b_i}{k}$  ou encore

$$\frac{\sum_{i \in P} a_i}{k} \leq \frac{\sum_{i \in P} b_i}{k} \text{ c'est-à-dire } m_a \leq m_b$$

*Proposition 2* :  $\Phi(-\tilde{q}_{(q)}(a, -b)) \geq \Phi(-\tilde{q}_{(q)}(b, -a))$  si et seulement si  $\text{Rang}(\text{Objet } B) \leq \text{Rang}(\text{Objet } A)$  dans le rangement synthétique.

Ceci signifie que le rangement de l'objet  $A$  à un degré de préférence donné «entraîne» le rangement de l'objet  $B$  à un degré de préférence plus élevé. Ces deux propositions permettent l'accord sémantique de l'indice avec l'intensité implicative où la variable  $A$  serait moins bien classée que la variable  $B$ .

Il advient alors que l'ordre des objets issu de l'ordre des intensités de propension permutationnelle est compatible avec celui issu de l'ordre des sommes des rangs attribués par les juges aux  $q$  objets dans le modèle du  $W$  de Kendall.

### 2.7. Apport de l'exploration des rangements des $q$ objets à l'aide de la relation de propension

Les propriétés précédentes laissent penser que finalement la relation de propension ne fait qu'apporter une information redondante à celle fournie par l'application de la procédure  $\bar{W}$  de Kendall. Toutefois elle nous paraît fournir une information plus analytique sur la structure de quasi-ordre affectée d'un niveau de confiance  $1 - \alpha$ , qui organise l'ensemble des  $q(q - 1)$  paires d'objets, que ne le permet l'approche par la procédure du  $\bar{W}$  de Kendall qui fournit un rangement global synthétique.

Rappelons les définitions posées par J.B. Lagrange (Lagrange 1998 p. 82) :

Un seuil de confiance  $1 - \alpha > 0$ , 50 étant choisi

- $A$  et  $B$  sont liées par la **relation de propension**

$$a\mathfrak{R}b \iff \begin{cases} \Phi(-\tilde{q}(a, -b)) \geq 1 - \alpha \\ \Phi(-\tilde{q}(a, -b)) \geq \Phi(-\tilde{q}(b, -a)) \end{cases}$$

- les  $n$ -uplets de variables distinctes  $(A_0, \dots, A_n)$  tels que

$$\left\{ \begin{array}{l} \forall i \quad \mathbf{A}_i \mathfrak{R} \mathbf{A}_{i+1} \\ \forall i \quad \forall j \quad \text{cov}(\mathbf{A}_i; \mathbf{A}_j) > 0 \\ \forall C \notin \{A_0, \dots, A_n\} \quad \text{non}(C\mathfrak{R}A_0) \text{ ou } \text{non}(A_n\mathfrak{R}C) \text{ ou } \exists i \quad \text{cov}(C, \mathbf{A}_i) \leq 0 \end{array} \right.$$

sont nommés **chemins de propension**.

- Une **filiation** est alors constituée par des chemins de propension ayant une extrémité commune.
- Deux variables  $A$  et  $B$  telles que  $\Phi(-\tilde{q}(a, -b)) = \Phi(-\tilde{q}(b, -a)) \geq 1 - \alpha$  sont dites **variables jumelles**.

Dans le contexte de l'étude des rangements des  $q$  objets, nous interprétons la propriété de covariance positive comme signifiant la tendance pour le rangement des deux objets  $A$  et  $B$  à recevoir des valeurs attribuées par les  $k$  juges qui vont dans le même sens.

Dans la procédure de Kendall, le  $q$ -uplet des moyennes empiriques des rangs attribués par les  $k$  juges aux objets  $(O_j)_{j=1, \dots, q}$  est l'information à partir de laquelle nous déterminons le rangement synthétique des  $q$  objets par les  $k$  juges. Il est clair que les valeurs de ces  $q$  moyennes dont la somme est constante, contiennent le niveau de préférence auquel chaque juge place chaque objet en relation avec les  $q - 1$  autres. Or dans la règle de rangement adoptée dans la procédure de Kendall, ce niveau n'est pas pris en compte. C'est alors que dans l'approche par la relation de propension permutationnelle, nous pouvons voir qu'il est rendu compte de ce phénomène. En effet si nous évaluons l'écart entre les deux coefficients de propension :

$$\begin{aligned}
\Delta \mathbf{P}_{(q)}^{(a,b)} &= \tilde{q}_{(q)}(a, \neg b) - \tilde{q}_{(q)}(b, \neg a) \\
\Delta \mathbf{P}_{(q)}^{(a,b)} &= \frac{\frac{\sum_{i \in P} a_i(1-b_i)}{k} - \frac{\sum_{i \in P} b_i(1-a_i)}{k}}{\sqrt{\frac{20q^4 - 46q^3 + 49q^2 - 41q + 24}{180k(q-1)^4}}} \\
&= \frac{\frac{\sum_{i \in P} a_i}{k} - \frac{\sum_{i \in P} b_i}{k}}{\sqrt{\frac{20q^4 - 46q^3 + 49q^2 - 41q + 24}{180k(q-1)^4}}}
\end{aligned}$$

Nous constatons que cet écart entre les deux valeurs empiriques du coefficient de propension est proportionnel à l'écart entre les deux moyennes empiriques. Plus l'écart entre les moyennes empiriques est fort, plus les valeurs empiriques des coefficients de propension de  $A$  vers  $B$  et de  $B$  vers  $A$  sont différenciées. De là alors que  $m_a \leq m_b$  permet de ranger l'objet  $B$  avant l'objet  $A$ , la quantité  $\Delta \mathbf{P}_{(q)}^{(a,b)} = \tilde{q}_{(q)}(a, \neg b) - \tilde{q}_{(q)}(b, \neg a) \leq 0$  complète la conclusion en fournissant un autre critère d'aide à l'interprétation basé sur une mesure, à savoir celle de l'intensité de la propension, et le niveau de confiance  $1 - \alpha$  dont elle rend compte.

## 2.8. Étude des situations extrêmes

### Concordance « parfaite »

La structure à explorer est celle d'un tableau dont les composantes de chaque colonne sont constantes et égales à une et une seule des valeurs de l'ensemble des degrés de préférence ou d'adhésion à un niveau de préférence  $\{0, \frac{1}{q-1}; \frac{2}{q-1}; \dots; \frac{q-2}{q-1}; 1\}$ . Cela signifie que tous les juges émettent le même jugement respectivement à chaque objet. Il en résulte que le vecteur  $(m_j)_{j=1, \dots, q}$  des moyennes empiriques des  $q$  colonnes est égal, à une permutation près, au vecteur  $(0, \frac{1}{q-1}; \frac{2}{q-1}; \dots; \frac{q-2}{q-1}; 1)$  de  $[0; 1]^q$ , tandis que le vecteur  $(\nu_j)_{j=1, \dots, q}$  des variances empiriques des  $q$  colonnes est égal au  $q$ -vecteur nul.

Quant à la quantité  $s(A \implies B) = \frac{\sum_{i \in P} a_i(1-b_i)}{k}$ , elle prendra l'une des valeurs suivantes issues des  $q(q-1)$  croisements possibles et qui coïncide trivialement avec  $m_a(1-m_b)$  :

TABLEAU 10

		Variable 1-B					
		Objets A ⇒ B	1	$\frac{q-2}{q-1}$	$\frac{q-3}{q-1}$	...	$\frac{1}{q-1}$
Variable A	0		0	0	0	0	0
	$\frac{1}{q-1}$	$\frac{1}{q-1}$		$\frac{q-3}{(q-1)^2}$	...	$\frac{1}{(q-1)^2}$	0
	$\frac{2}{q-1}$	$\frac{2}{q-1}$	$\frac{2(q-2)}{(q-1)^2}$		...	$\frac{2}{(q-1)^2}$	0
	...	...	...	...		...	0
	$\frac{q-2}{q-1}$	$\frac{q-2}{q-1}$	$\frac{(q-2)^2}{(q-1)^2}$	$\frac{(q-2)(q-3)}{(q-1)^2}$	...		0
	1	1	$\frac{q-2}{q-1}$	$\frac{q-3}{q-1}$	...	$\frac{1}{q-1}$	

Par exemple, si pour tout  $i \in \{1, 2, \dots, k\}$ , les réalisations respectives de  $A_i$  et  $B_i$  prennent respectivement les valeurs  $a_i = \frac{2}{q-1}$  et  $b_i = \frac{q-2}{q-1}$ , alors

$$s(A \implies B) = \frac{\sum_{i \in P} a_i(1-b_i)}{k} \text{ prend la valeur } \frac{2}{(q-1)^2}.$$

La valeur empirique du coefficient de propension dans le cadre de l'analyse implicative au sens de Lagrange devient  $\tilde{q}(a, -b) = \frac{s(A \implies B) - m_a(1 - m_b)}{m_a(1 - m_b)} = 0$  si  $m_a(1 - m_b) \neq 0$ . Dans la

sémantique de cette méthode d'analyse, cette situation où l'on ne peut se prononcer ni sur la propension de  $A$  vers  $B$ , ni sur sa réciproque correspond à une structure triviale de l'ensemble des rangs et, en fait, à un vide informationnel : si tous les juges sont d'accord, les préférences vont de soi, ce sont des données *a priori*.

Si  $m_a(1 - m_b) = 0$ , le rapport a une forme indéterminée  $\frac{0}{0}$ . Cette situation se produit quand l'objet  $A$  est au niveau de préférence 0 – le plus bas – pour tous les juges ou quand l'objet  $B$  est situé au niveau de préférence 1 – le plus haut – pour tous les juges. Replacé dans le contexte des rangements des  $q$  objets, cela signifie que l'objet  $B$  est l'objet préféré – donc  $\text{rang}(B) = 1$  – ou que l'objet  $A$  est celui qui est placé en dernier rang de préférence – donc  $\text{rang}(A) = q$  –. Il ressortait de l'approche par la statistique de Kendall que l'ensemble des objets possédait une structure d'ordre triviale significative. En analyse implicative, l'indétermination est levée par un passage à la limite, dans le cadre de la topologie de Fréchet. La valeur du coefficient de propension définie par continuité est nulle et correspond alors à une intensité de propension-implication statistique égale à 0,50. C'est aussi ce que vaut l'intensité d'implication dans le cas d'indépendance de deux variables binaires  $A$  et

$B$ . En ce qui concerne le facteur de variance, il est indéterminé quand  $m_a(1-m_a) = 0$  et infini dans les autres cas.

Si nous revenons au coefficient de propension intégrant l'approche permutationnelle, il devient :

$$\begin{aligned}\tilde{q}_{(q)}(a, -b) &= \frac{s(A \implies B) - \left[ \frac{1}{2} - \frac{(q-2)(3q-1)}{12(q-1)^2} \right]}{\sqrt{\frac{20q^4 - 46q^3 + 49q^2 - 41q + 24}{180k(q-1)^4}}} \\ &= \frac{m_a(1-m_b) - \left[ \frac{1}{4} + \frac{(q+1)}{12(q-1)^2} \right]}{\sqrt{\frac{20q^4 - 46q^3 + 49q^2 - 41q + 24}{180k(q-1)^4}}}\end{aligned}$$

Or, la propension de  $A$  vers  $B$  n'est sémantiquement envisageable que si  $0 \leq m_a \leq m_b \leq 1$ . De là  $0 \leq m_a(1-m_b) \leq m_b(1-m_b)$ . Comme  $m_b(1-m_b)$  est toujours au plus égal à  $\frac{1}{4}$ , alors le numérateur de  $\tilde{q}_{(q)}(a, -b)$  est toujours négatif.

En effet,  $\left[ \frac{1}{4} + \frac{(q+1)}{12(q-1)^2} \right]$  est supérieur à  $\frac{1}{4}$ . Ce résultat montre que, sur la base de la propension permutationnelle, il est possible de calculer la qualité numérique de cette propension qui reste significative d'une tendance de  $B$  à dominer généralement  $A$ . Il n'y a donc pas contradiction entre l'approche de Lagrange et l'approche permutationnelle, mais cette dernière est plus précise.

### *Hétérogénéité « parfaite »*

La structure à explorer est celle d'un tableau dont les composantes de chaque colonne sont telles que leurs sommes sont constantes ou presque à une unité près. Dans cette situation, le vecteur  $(m_j)_{j=1, \dots, q}$  des moyennes empiriques des  $q$  colonnes est soit égal au  $q$ -vecteur constant  $(\frac{1}{2}; \dots; \frac{1}{2})$  et  $m_a(1-m_b) = \frac{1}{4}$ . Soit il peut être égal, encore à une permutation près, au  $q$ -vecteur  $(\frac{1}{2} + \frac{1}{2k(q-1)}; \dots; \frac{1}{2} + \frac{1}{2k(q-1)}; \frac{1}{2} - \frac{1}{2k(q-1)}; \dots; \frac{1}{2} - \frac{1}{2k(q-1)})$  de  $[0; 1]^q$ .

Dans ce cas, nous en déduisons que  $m_a(1-m_b)$  prend nécessairement l'une des valeurs suivantes :

$m_a(1-m_b) =$		
$\frac{1}{4}$	$\frac{1}{4} - \left( \frac{1}{2k(q-1)} \right)^2$	$\frac{1}{4} \left( 1 - \frac{1}{k(q-1)} \right)^2$
$\frac{1}{4} \left( 1 + \frac{1}{k(q-1)} \right)^2$	$\frac{1}{4} \left( 1 + \frac{1}{k(q-1)} \right)$	$\frac{1}{4} \left( 1 - \frac{1}{k(q-1)} \right)$

Si nous considérons que le  $q$ -vecteur  $(m_j)_{j=1,\dots,q}$  des moyennes empiriques des  $q$  colonnes est approximativement égal au  $q$ -vecteur constant  $(\frac{1}{2}; \dots; \frac{1}{2})$  dans la mesure où  $k(q-1)$  est grand, la valeur empirique du coefficient de propension de J.B. Lagrange s'écrit maintenant :

$$\tilde{q}_{(q)}(a, -b) = \frac{\frac{\sum_{i \in P} a_i(1 - b_i)}{k} - \frac{1}{4}}{\sqrt{\frac{(\nu_a + \frac{1}{4})(\nu_b + \frac{1}{4})}{k}}}$$

### 2.9. Information apportée par la covariance : $cov(a; b)$

Dans l'approche de J.B. Lagrange, la valeur empirique du coefficient de propension peut s'exprimer en fonction de la valeur empirique de la covariance par la formule :

$$\tilde{q}(a, -b) = \frac{-cov(a; b)}{\sqrt{\frac{(\nu_a + m_a^2)(\nu_b + (1 - m_b)^2)}{k}}}$$

d'où la *proposition* :  $\Phi(-\tilde{q}(a, -b)) \geq 0,5 \iff cov(a; b) \geq 0$ . La valeur empirique  $cov(a, b)$  de la covariance de  $A$  et  $B$  contient une part importante de l'information relative à la relation de propension de  $A$  vers  $B$ . Dans notre approche permutatonnelle, la relation n'est plus aussi directe. En effet nous avons :

$$\begin{aligned} \tilde{q}_{(q)}(a, -b) &= \frac{\frac{\sum_{i \in P} a_i(1 - b_i)}{k} - \left[ \frac{1}{4} + \frac{(q+1)}{12(q-1)^2} \right]}{\sqrt{\frac{20q^4 - 46q^3 + 49q^2 - 41q + 24}{180k(q-1)^4}}} \\ &= \frac{m_a(1 - m_b) - cov(a; b) - \left[ \frac{1}{4} + \frac{(q+1)}{12(q-1)^2} \right]}{\sqrt{\frac{20q^4 - 46q^3 + 49q^2 - 41q + 24}{180k(q-1)^4}}} \end{aligned}$$

Si nous tenons compte du fait que les valeurs prises par  $A$  et  $B$  sont dans  $[0; 1]$ , nous pouvons déduire quelques informations relatives aux encadrements de diverses expressions :

$$0 \leq s(A \implies B) = \frac{\sum_{i \in P} a_i(1 - b_i)}{k} \leq 1$$

Par ailleurs nous savons que la suite  $M_q = \frac{1}{4} + \frac{(q+1)}{12(q-1)^2}$  est strictement décroissante sur  $IN^* - \{1\}$  et convergente avec pour limite  $M = \frac{1}{4}$

$M_2 =$	$M_3 =$	$M_4 =$	$M_5 =$	$M_6 =$	$M_7 =$	$M_8 =$	$M_9 =$	$M_{10} =$	$M_{11} =$
$\frac{1}{2}$	$\frac{1}{3}$	0,2962	0,2812	0,2733	0,26851	0,2653	0,2630	0,2613	0,26

En supposant qu'une étude non triviale requiert au moins trois objets à ranger, nous obtenons que  $\frac{1}{4} < \frac{1}{4} + \frac{(q+1)}{12(q-1)^2} \leq \frac{1}{3}$ . Dans cette situation où  $q > 2$ , la valeur empirique du coefficient de propension peut être ainsi encadrée

$$\sqrt{\frac{-\frac{1}{3}}{20q^4 - 46q^3 + 49q^2 - 41q + 24}} \leq \tilde{q}_{(q)}(a, -b) < \sqrt{\frac{\frac{3}{4}}{20q^4 - 46q^3 + 49q^2 - 41q + 24}}$$

Il advient aussi que la condition sur la valeur empirique de la covariance, à savoir  $\text{cov}(a; b) \geq 0$ , n'est plus une condition nécessaire et suffisante pour que le coefficient de propension soit négatif. En effet elle n'est plus nécessaire car dans l'exemple abordé dans cet article, nous rencontrons des cas où  $\text{cov}(a; b) < 0$  et  $\tilde{q}_{(q)}(a, -b) < 0$  – ainsi dans notre étude  $\text{cov}(V45E; V45C) < 0$  et  $\tilde{q}_{(q)}(a, -b) = -3,3491 < 0$  –. Elle n'est pas non plus suffisante puisque nous avons aussi  $\text{cov}(a; b) > 0$  et  $\tilde{q}_{(q)}(a, -b) > 0$  – ainsi dans notre étude  $\text{cov}(V45A; V45J) > 0$  et  $\tilde{q}_{(q)}(a, -b) = 1,4981 > 0$  –.

Nous considérons alors la condition  $\text{cov}(a; b) \geq 0$  comme une condition qui réduit la relation de propension aux couples d'objets tels que les variables  $A$  et  $B$  soient corrélées positivement, prenant en compte ainsi le fait que les valeurs attribuées aux deux objets vont dans le même sens. Par là même, elle se rapporte à la même classe de variables modales mise en œuvre dans la situation que celle déterminée dans l'approche de J.B. Lagrange.

### 3. Analyse statistique implicative comme outil d'exploration de la structure des rangements (cas du rangement incomplet et sans *ex aequo*)

Nous allons maintenant considérer le cas où il est demandé aux individus de procéder à un choix de  $t$  objets parmi les  $q$  proposés, puis de les trier par ordre décroissant de préférence. Nous pouvons considérer alors que le résultat est une réalisation du type du tableau 1, mais avec seulement  $t$  ( $t < q$ ) cases sur  $q$  comportant

une des permutations  $(1, 2, \dots, t)$ . Nous proposons alors la modélisation suivante qui nous ramène à une forme traitable selon l'approche de  $W$  de Kendall.

Les individus choisissent  $t$  objets et les rangent de 1 à  $t$ . Après quoi, nous postulons que cela reviendrait à dire que les  $(q - t)$  objets non choisis sont placés en *ex aequo* aux rangs  $t + 1$  à  $t + (q - t) = q$ . Nous leur attribuons alors le rang moyen  $\frac{1}{q-t} \sum_{j=1}^{j=q-t} (t+j) = \frac{t+1+q}{2}$ . Cette démarche permet de conserver la propriété d'égalité à  $\frac{q(q+1)}{2}$  de la somme des rangs par individu.

Nous appliquons alors la procédure du  $W$  de Kendall en prenant en compte l'existence de ces *ex aequo*. Cette prise en compte des *ex aequo* modifie la valeur  $S_{K_{\max}}$  obtenue dans le cas de la concordance parfaite entre les  $k$  juges qui fournissent tous le même rangement, c'est-à-dire à une permutation près  $(1, 2, 3, \dots, t, \frac{t+1+q}{2}, \dots, \frac{t+1+q}{2})$ . Il vient donc :

$$\begin{aligned} S_{K_{\max}} &= \sum_{j=1}^{j=t} \left( kj - \frac{k(q+1)}{2} \right)^2 - (q-t) \left[ \frac{k(q+1+t)}{2} - \frac{k(q+1)}{2} \right]^2 \\ &= k^2 \sum_{j=1}^{j=t} \left( j - \frac{(q+1)}{2} \right)^2 + \frac{k^2 t^2 (q-t)}{4} \end{aligned}$$

ou encore en développant et en réduisant  $S_{K_{\max}} = \frac{k^2}{12} [q^3 - (q-t)^3 - t]$ .

De là, nous obtenons la statistique

$$W^*_{(q,t)} = \frac{\sum_{j=1}^q \left( S(O_j) - k \frac{(q+1)}{2} \right)^2}{\frac{k^2}{12} [q^3 - (q-t)^3 - t]}$$

que nous utiliserons de la même façon que la statistique  $W$  de Kendall. Notons que si  $t = q - 1$  nous retrouvons le cas du rangement complet.

### 3.1. Variable modale associée au rangement de $t$ objets parmi $q$ réalisés par $k$ individus

Nous adoptons la transformation suivante en tenant le même raisonnement que dans le § 2.1.



TABLEAU 11  
Construction de la variable modale associée aux rangs

Rangs	Valeurs de la variable modale sur $[0; 1]$
1	1
2	$1 - \frac{1}{t} = \frac{t-1}{t}$
3	$1 - \frac{2}{t} = \frac{t-2}{t}$
...	...
$t$	$1 - \frac{t-1}{t} = \frac{1}{t}$
$t+1$	0
...	...
$t + (q-t) = q$	0

Ainsi la variable modale est définie par la relation suivante  $X_{i,j}^{(q,t)} = \frac{(t+1) - R_{i,j}^q}{t} I_{\{1,2,\dots,t\}}(R_{i,j}^q)$

### 3.2. Espérance et variance de $Z$

Nous tenons le même raisonnement que dans le § 2.4, en présentant quelques étapes de la démonstration.

$$E(Z) = \left[ \frac{t+1}{2q} \right] \left[ 1 - \frac{(t-1)(3t+2)}{6t(q-1)} \right]$$

Calcul de  $E[A_i(1 - B_i)]$

Posons  $Y = A_i(1 - B_i)$ , ainsi  $E(Y) = E(A_i) - E(A_i B_i)$ . Les variables modales  $A_i$  et  $B_i$  représentent les variables aléatoires d'attribution du degré d'importance par l'individu  $i$  à l'objet  $O_{j_1}$  et à l'objet  $O_{j_2}$ .

$E(A_i) = E(B_i) = \sum_{u=1}^{u=t} \frac{u}{t} + \frac{q-t}{q} \cdot 0 = \frac{t+1}{2q}$  et  $E(A_i B_i) = \frac{(t^2-1)(3t+2)}{12tq(q-1)}$   
nous permettent de déduire l'espérance de  $A_i(1 - B_i)$

$$E[[A_i(1 - B_i)]^2] = \frac{(t+1)[(t-1)(-40t^3 - 36t^2 - 5t - 6) + 30t^2(2t+1)(q-1)]}{180q(q-1)t^3}$$

Calcul de  $E[(A_i(1 - B_i))^2]$

Nous partons de  $Y^2 = [A_i(1 - B_i)]^2 = (A_i)^2 + (A_i B_i)^2 - 2(A_i^2 B_i)$ . En partant des distributions des variables  $A_i$ ,  $A_i B_i$  nous obtenons en désignant  $S_k^t$ , la somme des puissances  $k^{\text{ème}}$  des nombres entiers de 1 à  $t$ ,

$$E[(A_i)^2] = \frac{(t+1)(2t+1)}{6tq}, \quad \text{puis} \quad E[(A_i B_i)^2] = \frac{\left[ \left( \frac{S_2^t}{t^2} \right)^2 - \frac{S_4^t}{t^4} \right]}{q(q-1)}$$

et

$$E[(A_i^2 B_i)] = \frac{\left[ \frac{S_1^t S_2^t}{t} - \frac{S_3^t}{t^3} \right]}{q(q-1)}$$

d'où nous déduisons  $E[(A_i(1 - B_i))^2]$  en utilisant les expressions de  $S_k^t$  placées en annexe.

### 3.3. Expression du coefficient de propension pour des variables de rangs (cas du rangement incomplet sans ex æquo)

Dans ce contexte particulier, le coefficient de propension s'exprime alors de la façon suivante où  $t$  est le nombre d'objets à ranger parmi les  $q$  objets soumis au jugement de  $k$  individus :

$$\tilde{q}_{(q,t)}(a, \neg b) = \frac{\frac{\sum_{i \in P} a_i(1 - b_i)}{k} - \left[ \frac{t+1}{2q} \right] \left[ \frac{(t+1)(3t+2)}{1-6t(q-1)} \right]}{\sqrt{\frac{(t+1)[(t-1)(-40t^3 - 36t^2 - 5t - 6) + 30t^2(2t+1)(q-1)]}{180kq(q-1)t^3}}}$$

$$\tilde{q}_{(q,t)}(a, \neg b) = \frac{m_a(1 - m_b) - \text{cov}(a; b) - \left[ \frac{t+1}{2q} \right] \left[ \frac{(t+1)(3t+2)}{1-6t(q-1)} \right]}{\sqrt{\frac{(t+1)[(t-1)(-40t^3 - 36t^2 - 5t - 6) + 30t^2(2t+1)(q-1)]}{180kq(q-1)t^3}}}$$

Nous constatons que pour  $t = q - 1$ ,  $\tilde{q}_{(q,q-1)}(a, \neg b) = \tilde{q}_{(q)}(a, \neg b)$

#### 4. Application à l'exploration des représentations des étudiants à l'égard des objectifs d'enseignement de la statistique

Dans une recherche que J.C Régnier conduit sur la thématique de l'enseignement-apprentissage de la statistique dans un enseignement à distance (Régnier 2002), les représentations des étudiants relativement à la statistique ont été explorées, en particulier, à partir des rangements des objectifs déclarés de son enseignement. Durant l'année 2001-2002, la population des 125 étudiants concernés a été soumise à un questionnaire dont une question consistait à ranger les 10 objectifs de l'enseignement selon un ordre décroissant d'importance. De cette population nous avons extrait un échantillon d'observation de 73 individus dont nous traitons les réponses fournies relativement au rangement des 10 objectifs suivants :

TABLEAU 12

<i>Énoncés des objectifs</i>	<i>Rang</i>
V45A) expliciter les questions d'une problématique dont les réponses relèvent d'une approche statistique,	
V45B) décrire, traiter, analyser des données de manière pertinente dans le cadre d'une étude en particulier dans le domaine éducatif,	
V45C) faire le lien entre la réflexion analytique sur des questions relevant du champ de l'éducation, leur formalisation et leur traitement quantitatif,	
V45D) lire avec un regard critique et distancié, les conclusions de diverses études statistiques apparaissant dans des rapports de recherche en Sciences de l'Éducation,	
V45E) poursuivre de façon autonome et personnalisée un apprentissage en statistique afin d'enrichir ses acquis,	
V45F) poser un regard plus positif à l'égard d'un domaine largement exploité dans les <i>media</i> , dans le sens de ne pas considérer les résultats dans l'ordre du tout ou rien mais en les replaçant judicieusement dans leur domaine de validité,	
V45G) exploiter des notions et des démarches mathématiques à des fins d'outils, et de ce fait de modifier dans un sens positif le rapport souvent négatif que nombre entretient avec cette science,	
V45H) s'exercer à un raisonnement intégrant l'idée de « risque d'erreur » dans l'énoncé de ses conclusions.	
V45I) s'exercer à l'interprétation de phénomènes éducatifs sur la base de données statistiques sur des « faits éducatifs » et sur des relations entre ces « faits »	
V45J) s'exercer à la communication des résultats des analyses des données en distinguant clairement le modèle utilisé, de la réalité qu'il est supposé représenter, en séparant bien les traitements menés à l'intérieur du modèle, des interprétations reformulées dans le contexte du problème.	

Le calcul des sommes  $S(O_j)_{j=1,\dots,10}$  donne les résultats suivant :

TABLEAU 13

Objectifs	$O_1$ V45A	$O_2$ V45B	$O_3$ V45C	$O_4$ V45D	$O_5$ V45E	$O_6$ V45F	$O_7$ V45G	$O_8$ V45H	$O_9$ V45I	$O_{10}$ V45J
Sommes	322	276	338	349	552	408	459	500	392	419
Rangs issus de l'ordre des sommes	2	1	3	4	10	6	8	9	5	7

La mise en œuvre de la procédure du test  $W$  de Kendall fournit les informations suivantes : la réalisation empirique de la statistique  $W$  vaut  $w \approx 0,14775$  tandis que la valeur critique au seuil de  $\alpha = 0,01$  vaut  $w_c \approx 0,03298$ .

Elles conduisent à rejeter l'hypothèse  $H_0$  d'indépendance des rangements fournis par les étudiants et à admettre à un niveau  $\alpha = 0,01$  de risque de 1<sup>ère</sup> espèce qu'il existe une tendance à s'accorder sur l'ordre d'importance des objectifs :

$$(O_2; O_1; O_3; O_4; O_9; O_6; O_{10}; O_7; O_8; O_5)$$

$$(V45B; V45A; V45C; V45D; V45I; V45F; V45J; V45G; V45H; V45E)$$

#### 4.1. Exploration des 73 rangements des 10 objectifs par l'analyse de la propension statistique

Nous avons transformé le tableau des rangs par la conversion  $X_{ij}^{(q)} = \frac{q - R_{ij}^{(q)}}{q - 1}$  afin d'obtenir un tableau dont les valeurs sont les réalisations de variables modales dans  $[0; 1]$ . Nous avons ensuite soumis ce tableau de données à la procédure de l'analyse de la propension avec le logiciel CHIC 2.3 (Couturier 2001).

TABLEAU 14

Objectifs	$O_1$ V45A	$O_2$ V45B	$O_3$ V45C	$O_4$ V45D	$O_5$ V45E	$O_6$ V45F	$O_7$ V45G	$O_8$ V45H	$O_9$ V45I	$O_{10}$ V45J
Occurrence	45,333	50,444	43,556	42,333	19,778	35,778	30,111	25,556	37,556	34,556
Moyenne	0,621	0,691	0,597	0,580	0,271	0,490	0,412	0,350	0,514	0,473
Écart Type	0,264	0,256	0,254	0,308	0,298	0,333	0,342	0,271	0,269	0,334
Variance	0,070	0,065	0,064	0,095	0,089	0,111	0,117	0,073	0,073	0,111
Facteur de variance	3,3642	3,2665	3,7348	2,5700	2,2291	2,2537	2,0739	3,1046	3,4428	2,2400

Nous obtenons le graphe «implicatif» suivant :

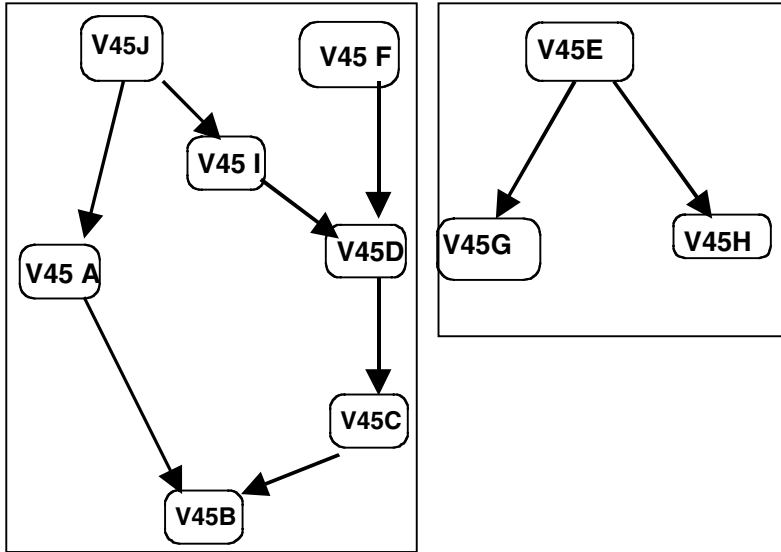


FIGURE 1  
*Graphe implicatif à partir de l'approche J.B. Lagrange de la relation de propension*

Les valeurs prises en compte pour construire ce graphe sont celles du tableau 15. Ici nous ne retenons que les cas où l'intensité de propension est supérieure à 0,50 avec lesquels nous procédons à l'identification des couples (A; B) vérifiant la

**relation de propension**  $a \mathfrak{A} b \iff \begin{cases} \Phi(-\tilde{q}(a, -b)) \geq 1 - \alpha = 0,5 \\ \Phi(-\tilde{q}(a, -b)) \geq \Phi(-\tilde{q}(b, -a)) \end{cases}$

TABLEAU 15

intensité de propension de X vers Y (J.B. Lagrange)

(X⇒Y)	V45A	V45B	V45C	V45D	V45E	V45F	V45G	V45H	V45I	V45J
V45A		0,7025								
V45B										
V45C		0,5816								
V45D			0,5825							
V45E							0,7953	0,6263		
V45F				0,6708						
V45G										
V45H										
V45I				0,5587						
V45J	0,5190								0,5837	

possédant la propriété de la covariance empirique positive. Le tableau 16 qui rapporte les valeurs empiriques des coefficients de corrélation linéaire, fournit aussi les couples dont la covariance empirique est positive.

TABLEAU 16

Corrélation (X ; Y)

(X⇒Y)	V45A	V45B	V45C	V45D	V45E	V45F	V45G	V45H	V45I	V45J
V45A	1,0000	<b>0,2490</b>	-0,1257	-0,2376	-0,1908	-0,2336	-0,1527	-0,0909	-0,0643	<b>0,0169</b>
V45B		1,0000	<b>0,0966</b>	-0,2052	-0,0200	-0,3369	-0,1382	-0,1794	-0,1598	-0,0777
V45C			1,0000	<b>0,0976</b>	-0,1115	-0,1142	-0,1349	-0,2241	-0,1336	-0,1838
V45D				1,0000	-0,3894	<b>0,1558</b>	-0,1822	-0,1429	<b>0,0631</b>	-0,2077
V45E					1,0000	-0,2323	<b>0,2596</b>	<b>0,1326</b>	-0,2534	-0,2185
V45F						1,0000	-0,0719	-0,3030	-0,0055	-0,0804
V45G							1,0000	-0,0107	-0,3412	-0,4027
V45H								1,0000	-0,1204	-0,0072
V45I									1,0000	<b>0,0885</b>
V45J										1,0000

Nous retrouvons dans ces tableaux, l'origine de 9 couples vérifiant la relation de propension au niveau de confiance minimal de  $1 - \alpha = 0,50$ .

Le relation de propension est compatible avec le rangement issu de la procédure  $W$  de Kendall. Mais elle s'avère plus informative car, d'une part, l'ordre est partiel et, d'autre part, les relations de préférences sont valuées

Objectifs	$O_2$	$O_1$	$O_3$	$O_4$	$O_9$	$O_6$	$O_{10}$	$O_7$	$O_8$	$O_5$
	V45B	V45A	V45C	V45D	V45I	V45F	V45J	V45G	V45H	V45E
Rangs	1	2	3	4	5	6	7	8	9	10

Voici les chemins de propension :

*Chemin JAB*

V45J [rang = 7]  $\implies$  V45A [rang = 2]  $\implies$  V45B [rang = 1]

*Chemin JIDCB*

V45J [rang = 7]  $\implies$  V45I [rang = 5]  $\implies$  V45D [rang = 4]  $\implies$   
V45C [rang = 3]  $\implies$  V45B [rang = 1]

*Chemin FDCB*

V45F [rang = 6]  $\implies$  V45D [rang = 4]  $\implies$  V45C [rang = 3]  $\implies$   
V45B [rang = 1]

*Chemin EG*

V45E [rang = 10]  $\implies$  V45G [rang = 8]

*Chemin EH*

V45E [rang = 10]  $\implies$  V45H [rang = 9]

Nous constatons le respect de la propriété de la propension à ce que «l'objet  $O_{j_1}$  étant rangé au rang  $r_1$ », «l'objet  $O_{j_2}$  est rangé au rang  $r_2$  qui traduit la préférence

de cet objet  $O_{j_2}$  sur l'objet  $O_{j_1}$  ». Nous n'abordons pas ici l'approche sémantique que requiert l'interprétation de ces résultats.

#### 4.2. Approche permutacionnelle de la propension de $X$ vers $Y$

Dans l'approche que nous avons nommée « permutacionnelle » en raison du fait que la base de l'information est une permutation sur  $\{1, \dots, q\}$ , nous suivons le même raisonnement. La seule différence réside dans le fait que le calcul de l'espérance et de la variance de la statistique  $Z$  utilisée par J.B. Lagrange tient compte de la particularité de l'information fondée sur la statistique de rangs. Nous utilisons les valeurs exactes vraies et non pas des estimations. Nous obtenons ainsi les résultats relatifs aux 90 couples de variables desquels nous extrayons les couples en relation de propension reportés dans le tableau 17 pour les niveaux de confiance  $1 - \alpha \geq 0,50$ . Pour calculer l'intensité de propension, nous utilisons encore la référence à une mesure gaussienne qui demeure monotone par rapport aux coefficients de propension permutacionnelle

TABLEAU 17  
Intensité de propension de  $X$  vers  $Y$  (approche permutacionnelle)  
Les valeurs marquées avec (+) correspondent aux couples de variables  
dont la covariance empirique est positive

( $X \Rightarrow Y$ )	V45A	V45B	V45C	V45D	V45E	V45F	V45G	V45H	V45I	V45J
V45A		0,97790+								
V45B										
V45C	0,73363	0,97388+								
V45D	0,69754	0,93808	0,7930+							
V45E	0,99959	0,99998	0,99959	0,99544		0,99024	0,99864+	0,98735+	0,99465	0,98811
V45F	0,90016	0,97087	0,89595	0,95205+					0,70302	
V45G	0,98327	0,99775	0,97388	0,94585					0,75517	
V45H	0,99780	0,99948	0,99270	0,99150		0,90222	0,89877		0,97291	0,96241
V45I	0,92510	0,98343	0,85127	0,8802+						
V45J	0,97412+	0,99428	0,89947	0,83124		0,60142			0,8211+	

La réalisation du graphe implicatif au seuil de confiance de 0,90 (et avec le critère de la covariance positive) conduit à la représentation des relations de propension suivante (Figure 2)

Si nous explicitons les chemins de propension nous obtenons alors :

*Chemin JAB*

V45J [rang = 7]  $\Rightarrow$  V45A [rang = 2]  $\Rightarrow$  V45B [rang = 1]

*Chemin CB*

V45C [rang = 3]  $\Rightarrow$  V45B [rang = 1]

*Chemin FD*

V45F [rang = 6]  $\Rightarrow$  V45D [rang = 4]

*Chemin EG*

V45E [rang = 10]  $\Rightarrow$  V45G [rang = 8]

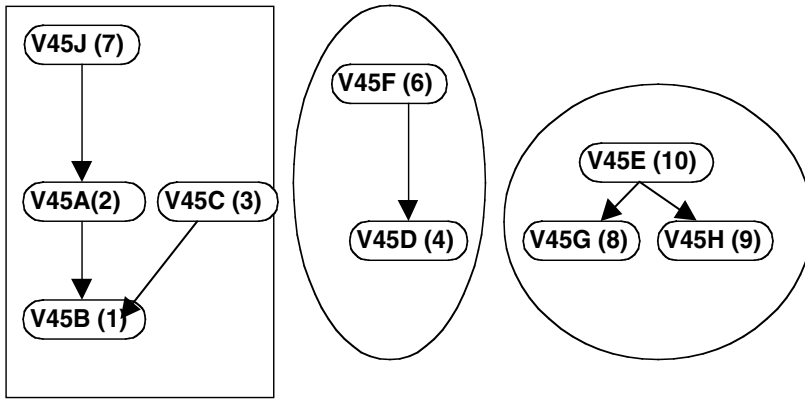


FIGURE 2

*Graphe implicatif à partir de l'approche permutacionnelle de la relation de propension*

*Chemin EH*

V45E [rang = 10]  $\implies$  V45H [rang = 9]

Au seuil de  $1 - \alpha = 0,50$ , nous obtenons en plus le *Chemin JIDCB*

V45J [rang = 7]  $\implies$  V45I [rang = 5]  $\implies$  V45D [rang = 4]  $\implies$   
V45C [rang = 3]

Comme nous l'avons démontré, cette structure définie par la relation de propension – *versus* approche permutacionnelle – respecte la propriété établissant que «l'objet  $O_{j_1}$  étant rangé au rang  $r_1$ », «l'objet  $O_{j_2}$  est rangé au rang  $r_2$  qui traduit la préférence de cet objet  $O_{j_2}$  sur l'objet  $O_{j_1}$ ».

### 4.3. Comparaison des résultats issus des deux approches

Dans le tableau 18, nous rapportons les valeurs des écarts des intensités de propension calculées respectivement dans le modèle dit approche permutacionnelle et dans le modèle dit de Lagrange.

Ce tableau met en évidence l'incidence sur la valeur de l'intensité de propension quand on ne tient pas compte des contraintes du modèle permutacionnel. En moyenne, l'intensité de propension permutacionnelle dépasse celle au sens de Lagrange, d'une valeur d'environ 0,13. Toutefois cette intensité de propension permutacionnelle reste inférieure à celle au sens de Lagrange, pour 40 couples sur 90, soit dans plus de 44 % des cas. Notons enfin que les propriétés topologiques du graphe implicatif, définies par les plus fortes intensités, sont peu affectées par ces différences, les principales liaisons étant donc respectées.



TABLEAU 18

Écart entre l'intensité de propension de  $X$  vers  $Y$  au sens de l'approche permutatonnelle et celle au sens de Lagrange (moyenne = 0,13451)

( $X \Rightarrow Y$ )	V45A	V45B	V45C	V45D	V45E	V45F	V45G	V45H	V45I	V45J
V45A		0,2754	0,1110	0,0105	-0,4046	-0,2961	-0,3955	-0,4532	-0,3103	-0,4450
V45B	-0,0159		-0,1631	-0,2133	-0,4910	-0,2899	-0,4183	-0,4190	-0,3856	-0,4455
V45C	0,3287	0,3923		0,0887	-0,4439	-0,3071	-0,4009	-0,3861	-0,2234	-0,3159
V45D	0,4046	0,6380	0,2106		-0,2777	-0,2997	-0,3460	-0,4114	-0,1857	-0,2645
V45E	0,7545	0,5321	0,6459	0,9225		0,7791	0,2033	0,3610	0,7760	0,7581
V45F	0,6397	0,8196	0,5109	0,2813	-0,3339		-0,2227	-0,2636	0,2081	0,0288
V45G	0,6664	0,6822	0,6263	0,6673	-0,3227	0,4258		-0,0628	0,5718	0,3629
V45H	0,6049	0,7245	0,7264	0,6620	0,0488	0,7090	0,4100		0,5929	0,4705
V45I	0,4831	0,6403	0,4623	0,3215	-0,3512	-0,0092	-0,2025	-0,3990		-0,0899
V45J	0,4552	0,5905	0,5845	0,5585	-0,3352	0,1871	-0,0877	-0,3580	0,2374	

### Conclusion

Nous avons rappelé que l'objectif du test de rang de Kendall consistait à établir un degré de concordance-discordance entre les jugements de  $k$  juges portant sur  $q$  objets, mais que sa référence à une loi de probabilité était délicate et que le résultat se limitait à rejeter l'hypothèse nulle. De même, dans le cas de rangement complet (tous les objets sont rangés) ou incomplet (certains peuvent ne pas l'être), la méthode, dite de propension, appliquée aux variables modales par J.B. Lagrange dans le cadre de la méthode implicative, ne pouvait être compatible avec la notion de rangement qui oblige à renoncer à la condition *a priori* d'indépendance entre les variables. Aussi, privés de cette propriété, nous avons établi les valeurs des paramètres des nouvelles variables en jeu dans une approche de la même propension exprimée de façon raccourcie par «l'objet  $b$  est généralement préféré à l'objet  $a$ ». Nous avons vu que la seule covariance, par son signe, n'était plus ni nécessaire ni suffisante pour conjecturer une relation de propension entre objets. Toutefois, une même référence à une mesure gaussienne pour déterminer l'intensité de propension dans les deux approches (permutatonnelle et celle de Lagrange) nous a permis alors de comparer, sur une application, les structures d'ordre partiel sur l'ensemble des objets pour les deux relations de propension. On a observé que les différences étaient faibles, particulièrement lorsque les intensités de propension étaient fortes, c'est-à-dire lorsque les relations de préférences étaient nettes. Une nouvelle question se pose alors : les différences entre les deux modèles sont-elles atténuées par l'admission des *ex aequo* où, cette fois, l'hypothèse d'indépendance entre les variables pourrait peut-être y trouver une nouvelle légitimité?

## Références

- BAILLEUL M. (1994), *Analyse statistique implicative : variables modales et contribution des sujets. Application à la modélisation de l'enseignant dans le système didactique*. Thèse Université de Rennes 1.
- COUTURIER R. (2001), Traitement de l'analyse statistique implicative dans CHIC, *Actes des Journées sur la Fouille dans les données par la méthode d'analyse implicative, IUFM Caen*, 33-50.
- FRIEDMAN M. (1937), The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. Statist. Ass.*, 32, 675-701.
- GRAS R. (1979), *Contribution à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs didactiques en mathématiques*, Thèse d'État, Université de Rennes 1.
- GRAS R. & COLL. (1996), *L'implication Statistique*, Grenoble, La Pensée Sauvage.
- GRAS R., KUNTZ p. , BRIAND H. (2001), Les fondements de l'analyse statistique implicative et leurs prolongements pour la fouille de données, *Mathématiques et Sciences Humaines*, n° 154-155, 9-29.
- KENDALL M.G., BABINGTON SMITH B. (1939), The problem of  $m$  rankings. *Ann. Math. Statist.*, 10, 275-287.
- LAGRANGE J.B. (1998), Analyse implicative d'un ensemble de variables numériques; application au traitement d'un questionnaire aux réponses modales ordonnées, *Revue de Statistique Appliquée XLVI-1*, Paris, 71-93.
- LECOUTRE J.-P., TASSI Ph. (1987), *Statistique non paramétrique et robustesse*, Paris Economica.
- LERMAN I.C., GRAS R., ROSTAM H. (1981), Élaboration et évaluation d'un indice d'implication pour des données binaires, I et II, *Mathématiques et Sciences Humaines*, 74, 5-35 et 75, 5-47.
- LERMAN I.C. (1981), *Classification et analyse ordinale des données*, Paris, Dunod,
- RÉGNIER J.C. (2002), Étude des difficultés d'apprentissage de la statistique dans le cadre d'un enseignement à distance. *Revue Éduquer : psychologie et sciences de l'éducation*, Paris L'Harmattan (47 p. ) (à paraître 2005).
- SAPORTA G. (1990), *Probabilités, analyse des données et statistique*. Paris, Éditions Technip.

## Annexe

**Rappels sur  $S_k^{(q)} = \sum_{r=1}^q r^k$ , la somme des puissances  $k^{\text{ème}}$  des  $q$  premiers nombres entiers :**

Par développement de  $\sum_{r=1}^q (1+r)^{k+1}$  on peut obtenir successivement

$k = 1$	$S_1^{(q)} = \sum_{r=1}^q r = \frac{q(q+1)}{2}$
$k = 2$	$S_2^{(q)} = \sum_{r=1}^q r^2 = \frac{q(q+1)(2q+1)}{6}$
$k = 3$	$S_3^{(q)} = \sum_{r=1}^q r^3 = \left[ \frac{q(q+1)}{2} \right]^2$
$k = 4$	$S_4^{(q)} = \sum_{r=1}^q r^4 = \frac{1}{30} [q(q+1)(2q+1)(3q^2+3q-1)]$

$S_1^{(q)} S_2^{(q)} - S_3^{(q)} = \frac{q^2(q+1)^2(q-1)}{6}$
$\left[ S_2^{(q)} \right]^2 - S_4^{(q)} = \frac{q(q-1)(q+1)(2q+1)(2q-1)(5q+6)}{180}$