LES CAHIERS DE L'ANALYSE DES DONNÉES

J.-P. BENZÉCRI

# Classification par ordre et comparaisons en paires

*Les cahiers de l'analyse des données*, tome 12, n° 4 (1987), p. 401-406

<http://www.numdam.org/item?id=CAD_1987__12_4_401_0>

*Les Cahiers de l'Analyse des Données*
*Vol XII - 1987 - n° 4 - pp.401-406*

# CLASSIFICATION PAR ORDRE ET COMPARAISONS EN PAIRES

# [ ORDRE PAIRE]

*J.-P. BENZECRI* *

A propos de l'article de P. Weingarden et S. Nishisato:

**Est-ce qu'une méthode de classification par ordre peut reproduire les résultats de comparaisons en paire?**

l'auteur écrit au Dr. Chuck Chakrapani, éditeur du C.J.M.R. (Journal Canadien de Recherche de Marché), en reprenant l'exemple considéré dans l'article.

## 0. Introduction

L'objet de l'article est de comparer deux formes de collecte des données utilisées pour déterminer, parmi un ensemble restreint d'objets, celui qui a la préférence du public. Les données peuvent résulter soit de classements de tout l'ensemble, ("classifications par ordre"), demandées aux sujets interrogés, soit de comparaisons des objets deux à deux, ("comparaisons en paire"). Nous ne reprendrons pas l'intéressante discussion des auteurs, (étant convaincu, d'autre part, qu'il vaut la peine de situer les préférences des sujets dans un large contexte), mais concentreront notre attention sur l'utilisation qu'ils font de l'analyse des correspondances pour extraire une structure ordinale d'un ensemble de comparaisons en paire.

Bien que le volume des données ne permette pas de mettre en valeur la diversité des traitements multidimensionnels, l'exemple nous paraît, en effet, bien choisi pour une introduction pédagogique à la méthodologie de l'analyse des correspondances.

## 1. Construction du tableau de correspondance

L'ensemble des objets de la présente étude est constitué de cinq marques, désignée chacune par un sigle de deux lettres. Une marque est caractérisée par

(*) Professeur de Statistique à l'Université Pierre et Marie Curie.

Les cahiers de l'analyse des données – 0339-3097/87/04 401 6/$ 2.60/ 

comparaison avec les quatre autres. Par raison de symétrie, il est naturel de considérer au même niveau les infériorités et les supériorités; et d'introduire des comparaisons fictives entre chaque marque et elle-même. D'où le tableau donné ci-dessous.

Comparaisons par paires entre cinq marques.

| 5 | HG | CF | GK | HC | CS |
|---|---|---|---|---|---|
| >HG | 50 | 40 | 44 | 35 | 16 |
| <HG | 50 | 60 | 56 | 65 | 84 |
| >CF | 60 | 50 | 47 | 38 | 8 |
| <CF | 40 | 50 | 53 | 62 | 92 |
| >GK | 56 | 53 | 50 | 39 | 16 |
| <GK | 44 | 47 | 50 | 61 | 84 |
| >HC | 65 | 62 | 61 | 50 | 14 |
| <HC | 35 | 38 | 39 | 50 | 86 |
| >CS | 84 | 92 | 84 | 86 | 50 |
| <CS | 16 | 8 | 16 | 14 | 50 |

On lit par exemple à l'intersection de la colonne GK et de la ligne <CF que, dans 53% des comparaisons effectuées entre GK et CF, la marque GK a été trouvée inférieure à CF (<CF). On notera que dans les comparaisons d'une marque XX avec elle-même, les deux taux, <XX et >XX ont été fixés à 50%.

Dans la suite l'ensemble des 10 lignes du tableau sera appelé: ensemble des qualités.

## 2. Résultats de l'analyse du tableau gobal

Nous considérerons la place des éléments sur l'axe 1; place qui apparaît à la fois sur le graphique plan (1,2) et sur le listage des facteurs.

Sur l'axe 1, les marques sont rangées de gauche à droite dans l'ordre naturel suivant:

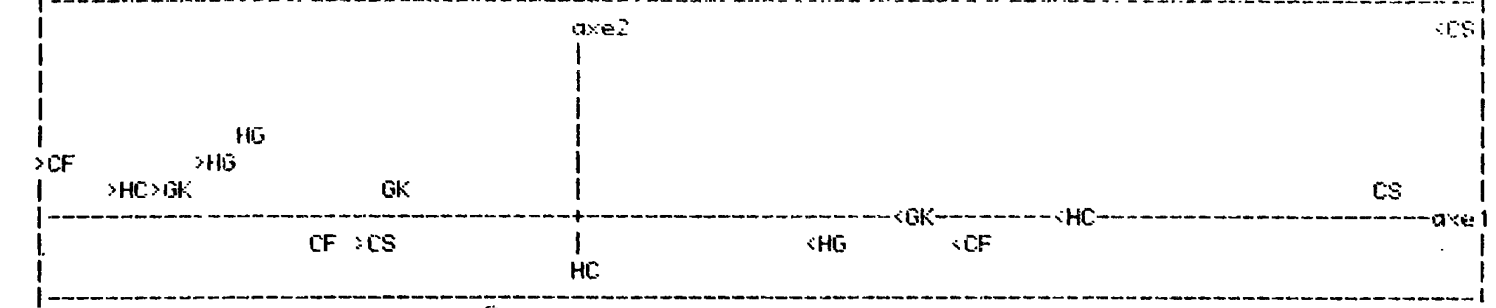
{HG, GF, GK, HC, CS},

c'est-à-dire que la marque HG qui est le plus généralement préférée à la marque CS qui est presque universellement rejetée.

Du côté (F1>0) on trouve avec CS toutes les qualités <XX (être jugée inférieure à XX); la position extrême étant celle de <CS elle-même dans les comparaisons fictives que nous avons introduites pour compléter le tableau des données.

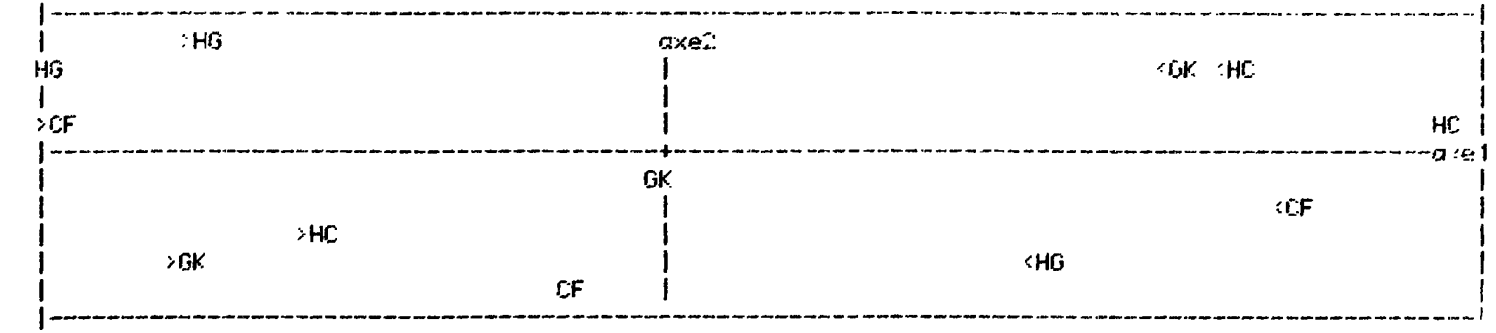
CAH marquesplantx : Remarques §2

ensemble(s) representa(s) j i
axe horizontal 1 , min=-4.31e-1 , max= 6.79e-1 , iam= 1.0be-1 ,taux= 9.60e-1
axe vertical 2 , min=-8.45e-2 , max= 2.26e-1 , iam= 3.97e-3 ,taux= 3.59e-2
element(s) j non represente(s) : 0
element(s) i non represente(s) : 0




vr marquesplantx : Remarques §4

ensemble(s) represente(s) : j i
axe horizontal 1 , min=-1.59e-1 ; max= 1.92e-1 ; iam= 1.38e-2 ;taux= 9.43e-1
axe vertical 2 , min=-4.10e-2 ; max= 4.04e-2 , iam= 6.74e-4 ;taux= 4.03e-2
element(s) j non represente(s) : 0
element(s) i non represente(s) : 0



Du côté (F1<0) on a, opposées à CS, les qualités>XX. On attendrait de voir en position extrême la qualité >HG: parce que, HG

étant la meilleure marque, lui être préférée est rare et singulier. Mais c'est >CF qu'on trouve; la raison est que (bien que CF vienne seulement au deuxième rang dans les préférences générales) CF est, de toutes les marques celle à laquelle la dernière marque (CS) est le moins souvent préférée. En effet, le plus faible nombre de la ligne <CS est inscrit dans la colonne CF.

Ainsi, au-delà des grandes lignes qui sont évidentes, il y a des détails qui nous encouragent à poursuivre l'étude.

### 3. Analyse avec la marque CS en supplémentaire

En analyse des correspondances, on appelle élément supplémentaire un élément (ligne ou colonne) qui (à la différence des éléments appelés principaux) ne participe pas à la création des axes; mais peut être projeté sur ceux-ci.

Ici, puisque la place de la marque CS est claire, on met CS en supplémentaire dans le but de voir mieux la configuration des quatre autres marques. Cependant les nuances notées au § 2 suggèrent de conserver comme éléments principaux les deux qualités <CS et >CS.

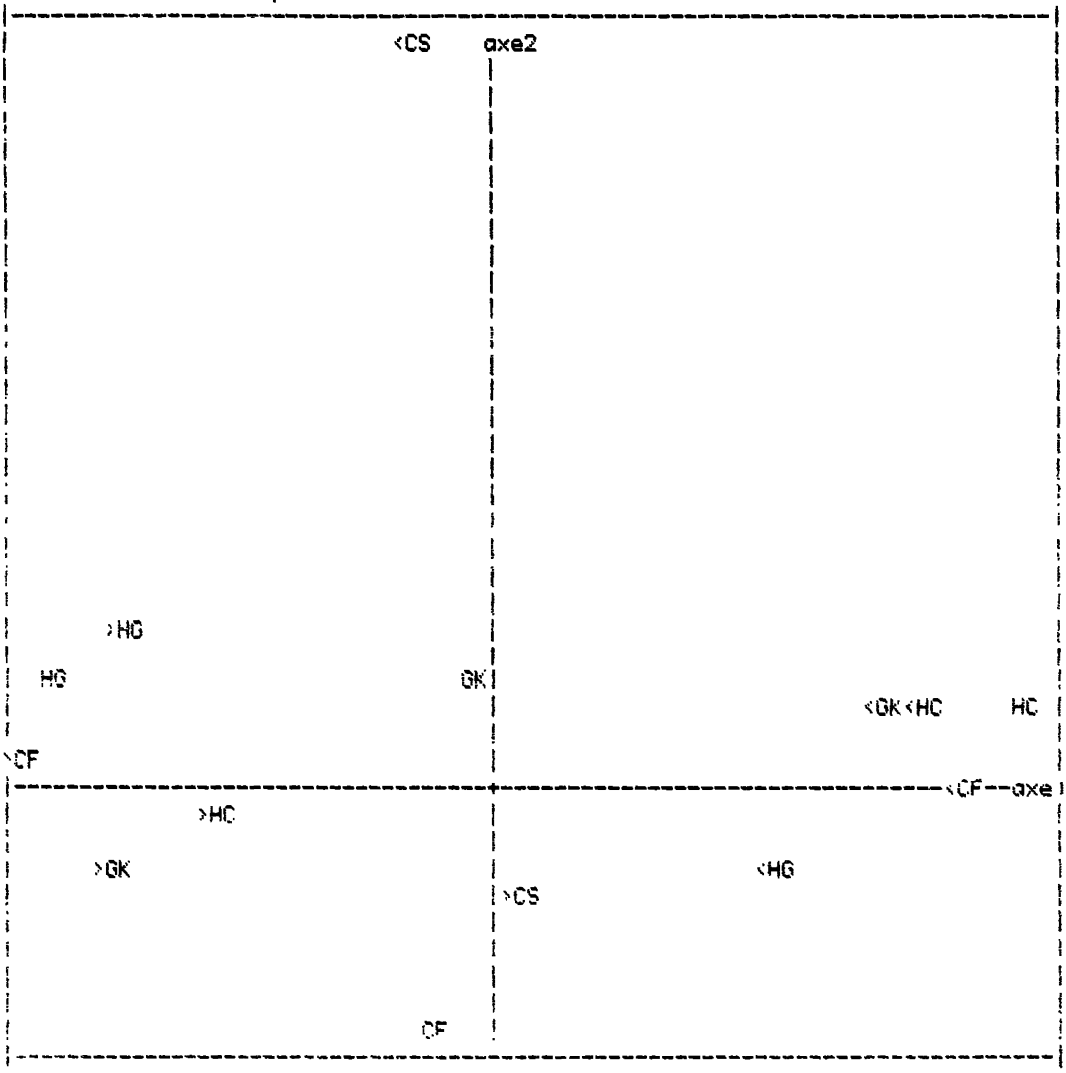
Dans le plan (1,2) les quatre marques {HG,GK,HC,CF}dessinent un triangle. A la marque HC, qui est la dernière de celles conservées, sont associées du côté (F1>0) les qualités <XX; (l'élément supplémentaire CS a sur l'axe 1positif une position tout à fait excentrique et c'est pourquoi on ne l'a pas mis sur le graphique plan). Tout à l'opposé, on a la meilleure marque HG associée aux qualités <XX; GK est intermédiaire entre HC et HG. Tandis que CF est au sommet inférieur du triangle, à l'extrémité négative de l'axe 2. Etant légèrement préféré à GK, CF se projette sur l'axe 1 un peu à gauche de GK; mais la principale différence entre ces deux marques s'inscrit sur l'axe 2, où CF s'oppose à <CS, parce que, comme on l'a remarqué, CF est, de toutes les marques, celle à laquelle CS est le moins souvent préférée.

Le triangle des quatre marques {HG,HK,HC,CF} se retrouve d'ailleurs sur le plan (1,2) de l'analyse globale (cf. § 2); mais comme il est orienté en travers des axes, l'interprétation y est plus difficile; et c'est pourquoi, au § 2, nous avons seulement considéré l'axe1.

### 4 Analyse restreinte à quatre marques et huit qualités

Si l'on met à la fois en supplémentaire la marque CS et les qualités <CS et >CS, plus rien ne subsiste que la sériation {HC,GK,CF,HG} sur l'axe 1. On le voit sur le graphique plan; et cette unidimensionnalité est confirmée par le fait que le taux d'inertie à l'axe 1 est de 94,3%.

```
CAH:vr:marquesplantx ; Remarques : §3
=======================================================================================
ensemble(s) represente(s) :      j     i
axe horizontal: 1 ; min=-1.59e-1 ; max= 1.72e-1 ; lam= 1.26e-2 ;taux= 8.30e-1
axe vertical  : 2 ; min=-8.24e-2 ; max= 2.41e-1 ; lam= 2.29e-3 ;taux= 1.50e-1
element(s)   j non represente(s) :     0
element(s)   i non represente(s) :     0
```




## 5 Conclusion

Répétons-le, ce n'est pas sur un petit ensemble de données que la statistique multidimensionnelle peut donner toute sa mesure; par exemple la classification automatique nous semble ici inutile. Mais avec seulement cinq marques on a pu

voir l'intérêt d'une approche progressive, qui met en relief des nuances; lesquelles dans une étude pratique peuvent avoir beaucoup plus d'intérêt que les grandes lignes évidentes *a priori*.. Par exemple, dans le cas présent, si l'on a interrogé plusieurs centaines de sujets, le fait que le minimum de la ligne <CS est dans la colonne CF est statistiquement significatif; et il vaut sans doute la peine d'en chercher une explication dans les goûts et attitudes des sujets.

N.B.: Les listages qui illustrent la présente lettre ont été réalisés sur ordinateur Macintosh par le programme qoriis, qui fait partie d'un ensemble conçu par l'auteur, comprenant classification, graphiques et aides à l'interprétation.