

Statistiques et informatique

appliquées aux sciences sociales

De l'Analyse d'un tableau croisé à l'analyse factorielle

Mardi 17/10/2023 15h30-17h D2 Censier

Licence de sciences sociales 3^e année
Université Paris 1 Panthéon Sorbonne

Plan de la séance

- Un exemple d'analyse des correspondances multiples (Cibois, 2002)
- Un exemple d'analyse factorielle (Desrosières, 1978)
- Récapitulatif des différentes méthodes
- Méthodes associées ou proches

Principes de l'analyse factorielle

- Analyser le tableau des écarts à l'indépendance
- Décomposer ces écarts comme la superposition d'effets « simples »
- Donner une représentation graphique (ici : un nuage de point) permettant de *projeter* chaque modalité en ligne ou en colonne sur des axes
- On représente les axes qui dans cette décomposition de l'information apportée par le tableau (mesurée par le Φ^2) en *expliquent* la plus grande part possible

Q3 4 Habituellement, quand prépares-tu ton cartable ? 1) Le matin, juste avant de partir au collège, 2) Le soir avant de dîner, 3) Le soir avant de t'endormir, 4) Tu n'as pas vraiment d'habitude fixe : cela dépend des jours. Sera considéré comme un indicateur de bonne volonté scolaire le fait de préparer son cartable la veille, donc de choisir les réponses 2 ou 3. Il n'est pas utile dans la suite de préciser la réponse correspondant à une forme de bonne volonté scolaire du fait de son évidence.

Q4 T'arrive-t-il d'oublier à la maison un livre ou un cahier ? 1) Souvent, 2) Parfois, 3) Rarement, 4) Jamais.

Q7 Penses-tu qu'arriver en retard au collège : 1) Ce n'est pas grave, cela arrive à tout le monde, 2) Ce n'est pas grave, si cela ne se produit que très rarement, 3) C'est grave et il faut l'éviter, 4) Tu ne sais pas.

Q8 Le plus souvent, les cours durent un peu moins d'une heure, trouves-tu que c'est : 1) Suffisant, 2) Trop long, 3) Trop court, 4) Tu ne sais pas.

Quand tu fais tes devoirs... 15A Ecoutes-tu la radio ? 1) Toujours, 2) Parfois, 3) Jamais.

15B Ecoutes-tu de la musique avec ton walkman ? 1) Toujours, 2) Parfois, 3) Jamais.

15C Allumes-tu la télévision ? 1) Toujours, 2) Parfois, 3) Jamais.

16 En général, quand tu as du travail scolaire à faire à la maison, quand le fais-tu ? 1) Le matin du jour où tu dois le rendre, 2) La veille du jour où tu dois le rendre, 3) Quelques jours avant de le remettre, 4) Le jour où il t'a été donné.

17 Habituellement, qui choisit le moment où tu dois travailler à la maison pour le collège ? 1) Tes parents ou un autre membre de la famille regardent ton cahier de textes et t'indiquent ce que tu dois faire, 2) Tu suis les conseils de tes professeurs, 3) Tu choisis tout seul le moment où tu commences à faire tes devoirs.

18 La quantité de travail (devoirs et leçons) qu'on te donne à faire en dehors de la classe est pour toi : 1) Importante, 2) Suffisante (ça peut aller), 3) Un peu insuffisante.

Quand travailles-tu à la maison pour le collège ?

19A Le soir en rentrant du collège 1) Toujours, 2) Souvent, 3) Rarement, 4) Jamais.

19B Le mercredi 1) Toujours, 2) Souvent, 3) Rarement, 4) Jamais.

19C Le samedi 1) Toujours, 2) Souvent, 3) Rarement, 4) Jamais.

19D Le dimanche 1) Toujours, 2) Souvent, 3) Rarement, 4) Jamais.

20 Lorsque tu dois aller en classe le lendemain, est-ce qu'il t'arrive de te coucher après 21 h 30 ? 1) Toujours, 2) Souvent, 3) Rarement, 4) Jamais.

Questions traitées en supplémentaires

NMA Appréciation du niveau de l'élève en mathématiques, 1) Bon, 2) Moyen, 3) Très passable, 4) Insuffisant.

SEX 1) Masculin, 2) Féminin

CSP du chef de famille 1) Agriculteurs, 2) Artisans, commerçants, chefs d'entreprises, 3) Cadres et professions intellectuelles supérieures, 4) Professions intermédiaires, 5) Employés, 6) Ouvriers, 7) Retraités, 8) Sans profession, 9) Chômage.

NAT 0) Français, 1) Maghreb, Turquie, Afrique noire, 2) Autre nationalité.

Prolongement : construction de variables idéaltypiques

« Pour voir l'effet spécifique, toutes choses égales par ailleurs, de [la] modalité « Maghreb », on forme avec les 5 modalités du type "bonne volonté scolaire" une variable Idéal-typique. En effet, un type mis en valeur par une analyse factorielle est en fait un idéal-type au sens wébérien, c'est à dire qu'il forme un type intelligible mais qu'il n'est pas ou peu attesté à l'état pur. Par exemple si l'on compte les individus de l'enquête qui ont zéro des 5 modalités du type, puis une, puis deux jusqu'à cinq (ceux qui auraient les 5 modalités du type, donc à l'état pur), on a la distribution suivante :

Nombre de modalités du type

Tot .	0	1	2	3	4	5
N=1865	402	662	556	197	44	4
%=100	21.6	35.5	29.8	10.6	2.4	0.2

On voit que le type à l'état pur ne comporte que 4 personnes, ce qui confirme que le type est un type-idéal. Si l'on accepte que pour faire partie du type, il suffit d'avoir plus que la moitié du nombre de modalités constitutives, c'est à dire ici 3, on voit que la population ayant ce type de bonne volonté scolaire regroupe 245 individus soit 13,1% de la population. C'est l'appartenance à cette sous-population qui va nous servir désormais de variable à expliquer dans une régression logistique. »

Exemple d'analyse factorielle

Marché matrimonial et structure des classes sociales.

Partant de l'hypothèse que les probabilités de mariages entre individus constituent une bonne mesure des distances entre leurs classes d'origine, on construit un modèle de l'espace social, tel que le révèlent ces probabilités. Le marché matrimonial est un des lieux où se négocient et se transmettent les divers types d'héritage dont disposent les agents : le mariage est un échange mettant en balance des espèces de capital variées, et la probabilité d'une union est d'autant plus forte que cet échange est plus «équilibré», et le mariage «bien assorti».

La statistique annuelle des mariages de l'Etat-civil, distribuant l'ensemble des 400 000 unions selon les catégories sociales croisées des pères de chacun des conjoints, permet d'une part d'analyser les échanges entre deux catégories, dont la fréquence est comparée à une fréquence fictive obtenue en supposant l'équiprobabilité des mariages entre classes, et d'autre part de confronter les destins matrimoniaux des fils (ou filles) de deux catégories, par l'analyse factorielle des correspondances de cette statistique : ces analyses montrent que les classes et fractions constituent un espace fortement structuré, tant par rapport au volume total de capital détenu, que du point de vue de la structure des diverses espèces de capital détenu (capital économique et capital culturel). Cette dernière différenciation est d'autant plus nette que le capital total est plus élevé : très marquée pour les classes supérieures, elle l'est moins pour les classes moyennes et est absente pour les classes populaires. L'homogamie sociale est forte, surtout pour les classes dominantes et pour les classes populaires, mais elle résulte, pour les premières, de mécanismes de sélection et d'aptitudes à maîtriser l'avenir, alors qu'elle reflète, pour les secondes, des mécanismes d'exclusion : l'examen des âges au mariage des hommes, de la fécondité ou des probabilités de divorce selon les milieux sociaux, confirme cette opposition entre des situations relativement choisies et des situations subies.

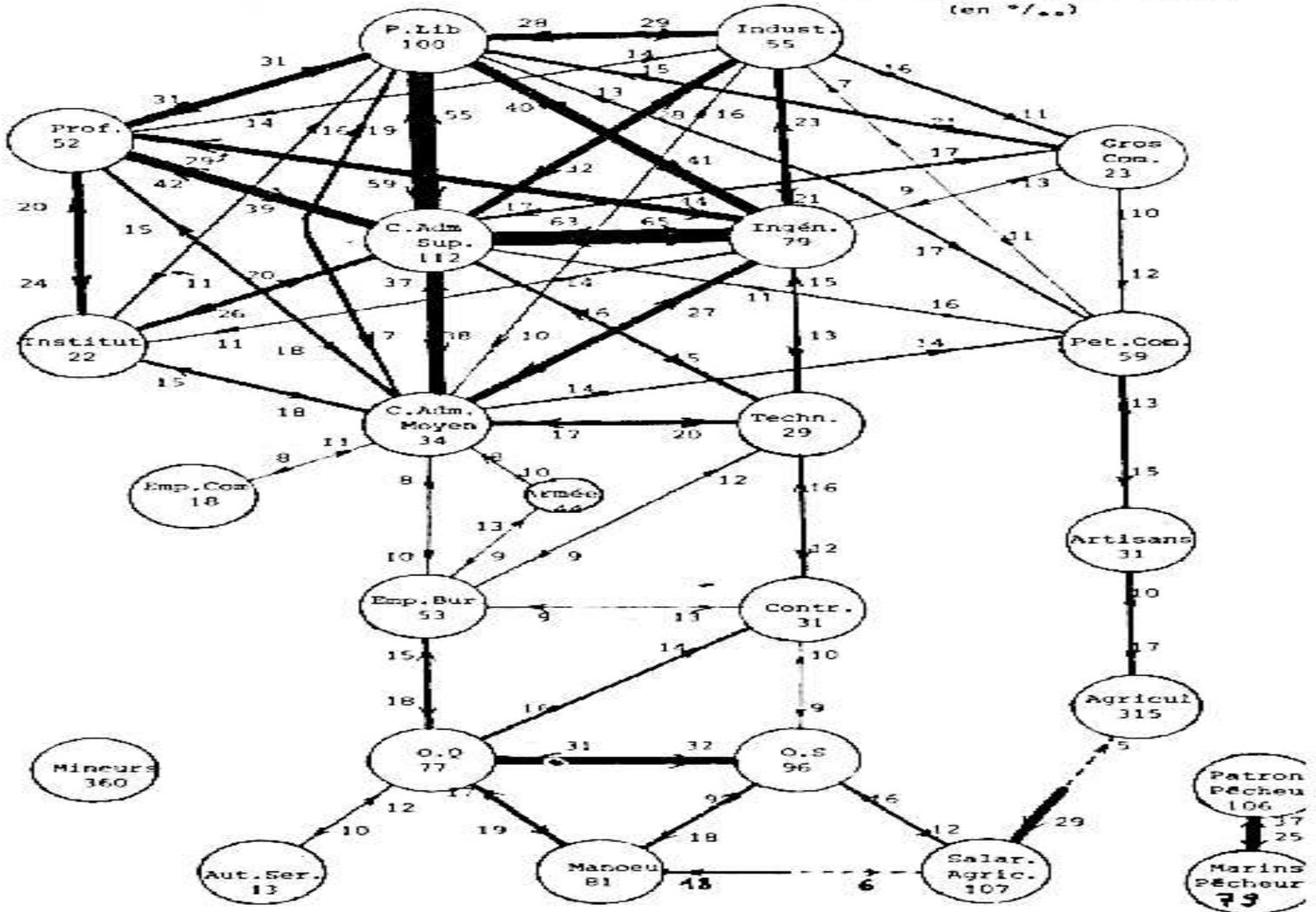
Alain Desrosières, 1978 (mars/avril), « Marché matrimonial et structure des classes sociales », *Actes de la recherche en sciences sociales*, 20-21 (Le patronat), 97-107;

Graphique I. Mariages et structure des classes sociales

$A \rightarrow B$ les fils de (A) épousent
 les filles de (B)

$$T_{ij} = \frac{P_{ij} - P_i \cdot P_j}{\sqrt{P_i \cdot (1 - P_i) \cdot P_j \cdot (1 - P_j)}}$$

(en %/100)



Exemple d'analyse factorielle

Alain Desrosières, 1978 (mars/avril), « Marché matrimonial et structure des classes sociales », *Actes de la recherche en sciences sociales*, 20-21 (Le patronat), 97-107;

Graphique II. Les origines sociales des conjoints (axe 1 x axe 2)

21 CS du père du mari
 X 21 CS du père de la femme
 (Etat-Civil 1972)

P. Lib.
 P. Lib.

Indus.
 Indus.

CAPITAL ECONOMIQUE

Ing.
 Ing.
 Prof. C.Adm.Sup.
 Prof. C.Adm.Sup.

Gros Com.
 Gros Com.
 Artistes

CAPITAL CULTUREL

Artistes
 C.Adm.Moy.
 C.Adm.Moy.
 Instit.
 Instit.

Pet. Com.
 Pet. Com.
 Technic.
 Technic.

SMed.
 Artisans
 Emp.Com.
 Emp.Com.
 Armée
 Armée

Autr.Serv.
 Contr.Autr.Serv.
 Contr.
 Emp.Bur.
 Emp.Bur.

O.Q.
 O.Q.

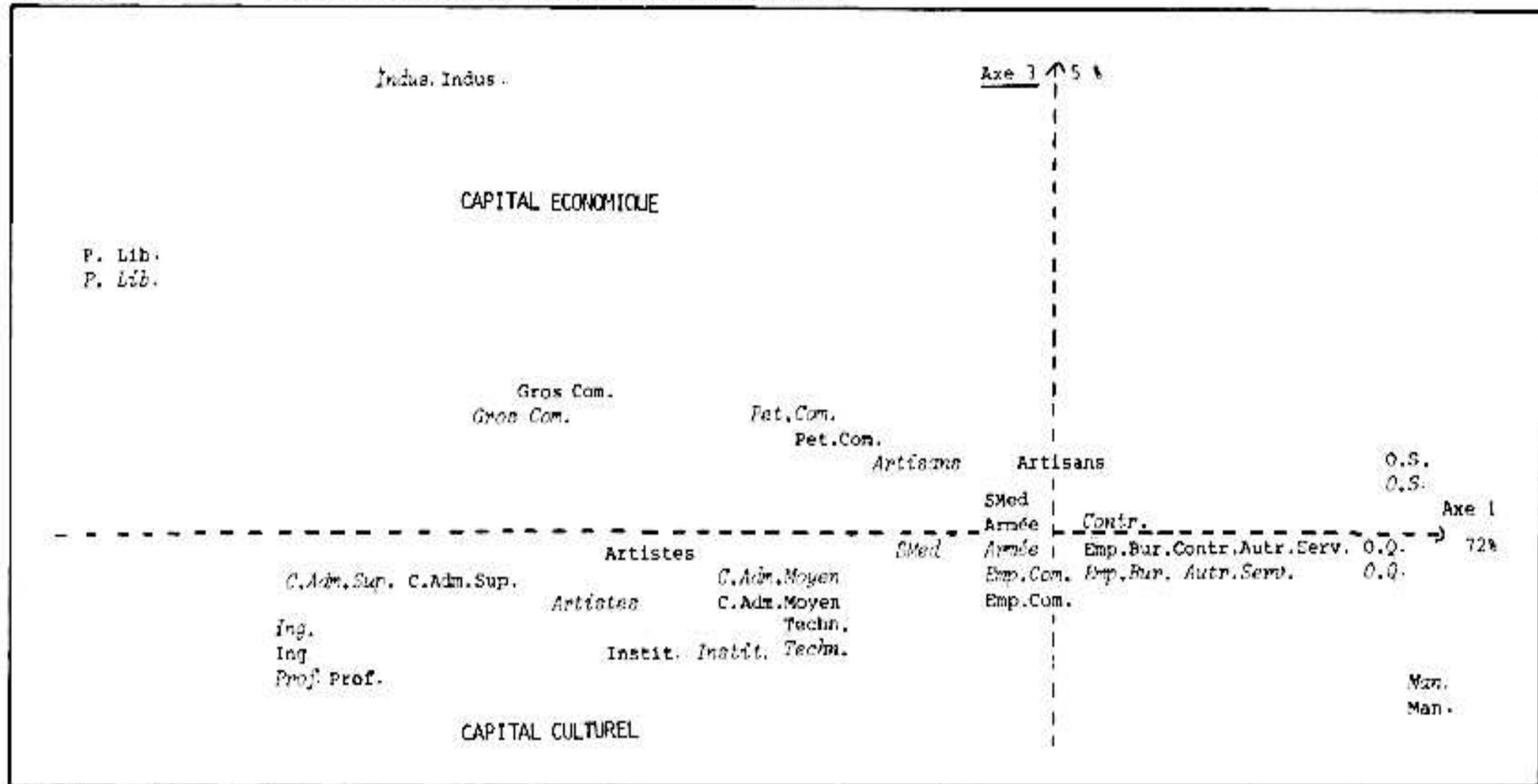
O.S.
 O.S.

Axe 1
 72

Axe 2

Man.
 Man.

Graphique III. Les origines sociales des conjoints (axe 1 x axe 3)



- Le graphique factoriel reflète les proximités/distances entre structures d'alliance d'une origine sociale à l'autre
- Élimination de catégories peu nombreuses et très homogames (déterminent sinon les 1^{ers} axes)
- Premier axe : 72 % de l'inertie, oppose les professions libérales aux manœuvres (selon la hiérarchie des CSP)
- Deuxième axe : oppose les catégories moyennes aux catégories élevées
- Troisième axe (5%): fait ressortir la structure des ressources détenues (capital économique vs culturel), déjà partiellement visible dans le premier graphique
- Pas d'asymétries détectables entre mariages des filles et des garçons
- Effets des mobilités sociales (tire « vers le haut » les enfants d'instituteurs par exemple)

Récapitulatif des différentes méthodes

- **Analyse des correspondances multiples (ACM)**

Sur de multiples variables qualitatives

AFC réalisée sur un tableau disjonctif complet ou un tableau de Burt (méthodes équivalentes)

Inertie du nuage et décomposition non interprétables

$$\frac{NTM - NTV}{NTV}$$

→ variables explicatives en supplémentaires, projetées

sur le nuage factoriel des variables d'intérêt

→ on s'arrête à deux axes ou juste avant un « saut »

- **Analyse factorielle des correspondances (AFC)**

Sur 2 variables qualitatives, éventuellement avec beaucoup de modalités (grand tableau)

L'inertie du nuage est égale au ϕ^2 du tableau croisé correspondant

L'AFC est un cas particulier d'ACP réalisée avec la métrique du Khi^2 pour calculer une distance entre deux profils lignes ou deux profils colonnes

Pas d'asymétrie entre variables

- **Analyse en composantes principales (ACP)**

Sur plusieurs variables quantitatives. Méthode de référence. Utilisable (en psychologie)

sur des modalités de variables qualitatives (avec variables dichotomiques centrées et réduites)

Méthodes associées ou proches

- Analyse textuelle (repose souvent sur un codage puis une analyse factorielle sous-jacente... mais plus sophistiquée)
- Méthodes de classification (CAH, CDH) pour découper des catégories et délimiter des groupes homogènes
- Remarque : il est possible de réaliser une analyse factorielle sur un tableau comportant de multiples variables en lignes et en colonnes (en démultipliant l'effectif)...
- ... on peut également si l'effectif le permet faire une analyse factorielle entre croisements de variables en ligne et en colonne (par exemple sexe et âge ou sexe et PCS en ligne, etc.)