

## Un dialogue de sourds dans le théâtre statistique ?

### Analyse géométrique des données et effets de structure<sup>1</sup>

Xavier Bry (I3M, Université de Montpellier II)

Nicolas Robette (Printemps, CNRS-UVSQ)

Olivier Roueff (CRESPPA-CSU, CNRS)

#### Les usages de la régression

Depuis leur introduction à la fin des années 1960, les modèles de régression se sont largement diffusés dans les sciences sociales anglo-saxonnes, pour rapidement devenir l'approche quantitative dominante, dans la continuité (entre autres) des travaux de Paul Lazarsfeld. Étienne Ollion a ainsi montré que les régressions n'étaient présentes que dans environ 10% des articles de la revue *American Sociological Review* entre 1967 et 1970, mais que cette proportion approchait les 70% à la fin des années 1990 (Ollion, 2011).

Cette « sociologie des variables » a fait l'objet d'un grand nombre de critiques, des plus générales aux plus techniques, des ethnométhodologues jusqu'aux statisticiens. Dans la seconde partie des années 1980, des sociologues étatsuniens ayant une certaine autorité dans le champ académique, comme Andrew Abbott (2001), Charles Ragin (1987) ou Stanley Lieberman (1985, voir aussi Vallet 2004), attaquent la sociologie des variables sur le terrain des hypothèses qui la sous-tendent.

Tout d'abord, cette « réalité linéaire générale », pour reprendre les termes d'Abbott, en cherchant à expliquer une caractéristique par des facteurs causaux, fait l'hypothèse que ces facteurs causaux agissent indépendamment les uns des autres (ce que Ragin appelle le « net effects thinking »). Ils ont un effet propre et, par conséquent, ne se conjuguent pas mais s'additionnent<sup>2</sup>. On pourrait ainsi comprendre le risque d'être touché par la pauvreté en séparant l'effet de l'origine sociale de celui du capital scolaire : le fait d'être sans diplôme aurait les mêmes conséquences pour un enfant d'ouvrier que pour un enfant de chef de grande entreprise.

De même, ces facteurs causaux, autrement dit les propriétés sociales des individus, ont une signification unique (ils sont univoques) et s'appliquent uniformément en tout point de l'espace social, quel que soit le contexte. Par exemple, un indicateur de pratique culturelle comme le fait de sortir au cinéma une fois par mois aura la même signification à Paris ou dans des régions rurales, pour les générations nées après-guerre comme pour celles de la seconde massification scolaire, etc<sup>3</sup>.

De plus, les cas (individus, pays, etc.) étudiés sont – eux aussi – supposés indépendants les uns des autres : ce qui arrive à un cas n'a pas d'influence sur ce qui arrive aux autres. Or en s'intéressant – pour différents pays – à l'impact de l'entrée en guerre sur la fécondité durant la Seconde Guerre mondiale, Lieberman montre que l'entrée en guerre de la Norvège y a modifié les comportements de fécondité, mais que la fécondité a aussi changé en Suède, pays pourtant resté neutre : c'est ce qu'il appelle un effet de contamination (ou de diffusion). Ici, le facteur causal peut avoir un effet même dans les cas où il n'est pas

---

<sup>1</sup> Les auteurs remercient chaleureusement Jérôme Deauvieux et Frédéric Lebaron pour leurs relectures, les discussions et les conseils avisés.

<sup>2</sup> Il est techniquement possible d'ajouter des effets d'interaction entre variables explicatives dans un modèle de régression mais, outre que c'est une pratique relativement peu répandue, la multiplication des croisements devient rapidement inexploitable.

<sup>3</sup> On trouve déjà la critique de ce postulat chez Halbwichs dans *Les causes du suicide* (2002 [1930]), à propos du lien entre religion et urbanisation.

directement présent. C'est plus généralement la question des interactions sociales qui est évacuée par l'hypothèse d'indépendance des cas.

Par ailleurs, on raisonne souvent « toute chose égale par ailleurs ». Or les sciences sociales ne peuvent jamais reproduire les dispositifs expérimentaux des sciences de la nature (population témoin, etc.)<sup>4</sup>. Il reste toujours des propriétés que l'on n'a pas observées, parce que l'on ne dispose pas de l'information ou parce qu'elles ne sont pas réductibles à un simple indicateur : la procédure de contrôle est toujours incomplète.

Enfin, les modèles de régression supposent généralement la réversibilité des processus causaux : si un facteur a une conséquence donnée, la disparition du facteur entraînera le retour à la situation initiale. Or on ignore ainsi que certains phénomènes sociaux sont difficilement réversibles, du fait par exemple d'effets cliquet<sup>5</sup>.

Au final, ces diverses attaques n'ont eu que peu d'effet sur la domination de l'approche par les variables dans les sciences sociales anglo-saxonnes. Le paradigme dominant a en fait contre-attaqué en proposant des modèles statistiques toujours plus complexes. Ces raffinements techniques sont supposés lever certaines objections, mais ne remettent jamais vraiment en question les fondements épistémologiques de l'approche.

Quels sont alors les usages pratiques de la régression et leurs présupposés théoriques, qui restent souvent à l'état d'impensés ? A grands traits, on peut distinguer trois types d'usages des modèles de régression. Le premier, un usage « mesuré », consiste à contrôler la structure de la population étudiée, i.e. à raisonner « à propriétés sociales égales par ailleurs ». L'article de Louis-André Vallet et Jean-Paul Caille sur la réussite scolaire des enfants d'immigrés (1995) en constitue un exemple classique. Les auteurs partent du constat que les enfants d'immigrés sont relativement moins nombreux à obtenir une orientation en seconde après quatre années de collège. Mais ils montrent ensuite qu'à caractéristiques familiales égales (profession du chef de famille, diplôme des parents, etc.), la différence de réussite disparaît : il n'y a pas d'effet propre de l'origine immigrée. C'est dans ce premier type d'usage qu'on peut situer le Durkheim du *Suicide*, même si ses « variations concomitantes » ne bénéficient pas encore des techniques de régression proprement dites.

Le second usage, qu'on pourrait qualifier de « métrologique » (Desrosières, 2001), consiste à utiliser la régression pour tenter de hiérarchiser l'importance de différents facteurs causaux. On tente par exemple de déterminer si c'est l'origine sociale ou le capital scolaire (l'hérité ou l'acquis) qui a le plus d'influence sur les goûts et les styles de vie ; ou de déterminer si ce sont les normes culturelles ou le milieu social qui expliquent l'intégration (ou la non-intégration) des immigrés<sup>6</sup>.

Le troisième usage, « hyper-métrologique », est plus directement lié aux procédures d'évaluation de l'action publique. Comme le montre Étienne Pénissat (2011) dans le cas des politiques d'emploi, il s'agit de mesurer le plus précisément possible l'efficacité, i.e. l'effet « pur », des dispositifs de retour à l'emploi, en calquant la démarche sur celle des sciences expérimentales : les outils statistiques participent ici à la redéfinition de l'action du champ bureaucratique<sup>7</sup>.

## **Analyse des données et régression : quelle complémentarité ?**

---

<sup>4</sup> Cet argument est aussi très présent dans les réflexions épistémologiques de Passeron (1991).

<sup>5</sup> On ne cite ici que quelques-unes des principales critiques, en laissant notamment de côté celles qu'Abbott consacre aux questions de temporalité (2001).

<sup>6</sup> On peut se demander si l'on ne comprendrait pas mieux certains débats scientifiques en analysant ces sortes de concours sur la « taille » et la « robustesse » des effets comme une forme de compétition virile.

<sup>7</sup> C'est la même logique qui explique le succès récent, auprès des organisations internationales étatiques comme non gouvernementales, de l'« évaluation aléatoire » promue par Esther Duflo (2010) pour tester l'efficacité des programmes d'aide au développement.

En France, la situation est bien différente de celle des pays anglo-saxons. Tout d'abord parce que les approches quantitatives en général y sont moins répandues. En ce qui concerne plus spécifiquement les régressions, on voit apparaître quelques appropriations diffuses à partir des années 80, notamment dans les travaux sur la stratification et la mobilité sociale à partir des grandes enquêtes de l'INSEE, mais on est encore loin d'un phénomène massif.

Par ailleurs, la trajectoire de Pierre Bourdieu n'est sans doute pas pour rien dans cette spécificité française, en particulier la forte association entre son programme de recherche et les méthodes d'analyse des correspondances (Rouanet *et al*, 2000 ; Lebaron, 2010). En 1966, dans un chapitre de l'ouvrage « Le partage des bénéfiques » intitulé « La fin d'un malthusianisme », Pierre Bourdieu et Alain Darbel (statisticien à l'INSEE) proposent une équation mathématique de la fécondité en fonction du niveau de revenus, en s'inspirant des modèles économétriques sur la consommation, et apportent ainsi des éléments d'explication sur le Baby-Boom que connaît alors la France. Mais la modélisation de la fécondité en fonction des revenus est immédiatement suivie d'une violente attaque contre la sociologie des variables, et sa volonté de séparer les influences respectives des déterminants sociaux de la fécondité, quand c'est plutôt la « structure de la constellation de facteurs » qu'il convient de décrire, et la manière dont cette structure se réfracte dans un groupe social donné.

Quelques années plus tard, en 1976, Bourdieu a – temporairement – formalisé sa théorie des champs, de nouveaux outils statistiques sont disponibles (en particulier grâce aux travaux de Benzécri), et il publie avec Monique de Saint-Martin « L'anatomie du goût », sa première application de l'analyse des correspondances à son modèle théorique. Cette application sera suivie de nombreuses autres, jusqu'aux « Structures sociales de l'économie » en 2000, et Bourdieu soulignera à plusieurs reprises l'affinité entre l'analyse des correspondances – ou analyse géométrique des données – et son approche relationnelle du monde social (Lebaron, 2010).

Dans ce contexte, l'articulation entre analyse géométrique et modèles de régression fait l'objet de débats, qui peuvent tourner au « dialogue de sourds » (Desrosières, 2003). Plutôt que d'opposer les démarches, on souligne parfois leur complémentarité (des Nétumières, 1997) : la division du travail statistique associe à l'analyse des correspondances l'exploration et la description des données, quand les régressions viennent expliquer, infirmer ou confirmer des hypothèses. Mais cette dichotomie est pour le moins simpliste : bien utilisée, l'analyse des correspondances « n'a rien de la méthode purement descriptive que veulent y voir ceux qui l'opposent à l'analyse de régression » (Bourdieu, 2000). Plus généralement, « la statistique n'explique rien, mais elle donne des éléments possibles d'explication » (Lebart *et al*, 2000). L'idée d'une complémentarité n'est donc pas acceptable en ces termes et doit être reformulée : Rouanet et Lebaron (2006) proposent ainsi de « chercher une synthèse entre analyse des données et régression, visant à intégrer la régression dans l'analyse géométrique, et à donner un sens à la sociologie des variables, replacée dans le cadre de l'espace social ». Autrement dit, il s'agit de conserver le groupe social ou l'individu comme acteur principal du « théâtre statistique » (Desrosières, 2003), et donc d'utiliser la régression (dans le premier de ses usages identifiés plus haut), sans adopter le « langage des variables » (Desrosières, 2001).

Dans la suite de cet article, nous décrivons une méthode d'analyse géométrique des données qui, dans la continuité des travaux de Rouanet *et al* (2002), se propose « non pas d'éliminer les effets de structure », en courant le risque notoirement souligné par Simiand avec sa métaphore du renne au Sahara, « mais de les analyser ».

## **L'espace des goûts cinématographiques**

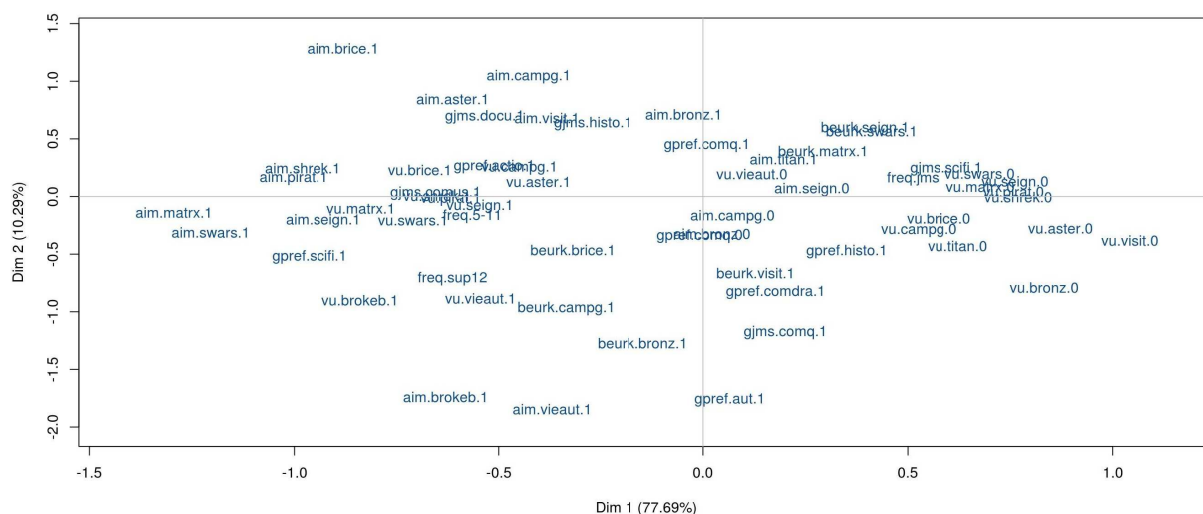
Nous nous intéressons ici à l'espace des goûts cinématographiques, construit à partir de l'enquête sur les *Pratiques Culturelles des Français* réalisée en 2008 par le Département des Études de la prospective et des

Statistiques du ministère de la Culture (Donnat, 2009)<sup>8</sup>. Les variables retenues pour la construction de cet espace décrivent à la fois des pratiques de fréquentation, des préférences positives *in abstracto* (goûts) et des préférences négatives *in abstracto* (dégoûts) :

- les genres de films préférés<sup>9</sup> ;
- les genres de films jamais regardés<sup>10</sup> ;
- les films que l'on a déjà vus parmi une liste (*Les Bronzés, Pirates des Caraïbes, Camping, Star Wars (La Guerre des étoiles), Brice de Nice, Shrek, Matrix, Titanic, Le Seigneur des anneaux, Astérix et Obélix, Les Visiteurs, La Vie des autres, Le Secret de Brokeback Mountain*) ;
- parmi ces films, ceux que l'on aime particulièrement ;
- parmi ces films, ceux que l'on n'aime pas du tout ou qu'on n'a pas du tout envie de voir ;
- la fréquence des sorties au cinéma.

On réalise une Analyse des Correspondances Multiples (ACM) à partir de ces variables, en se concentrant sur la population active (N=3008)<sup>11</sup>. Le premier axe de l'ACM peut être interprété comme une dimension d'intensité de la pratique cinématographique. On trouve notamment sur la partie gauche du graphique (Graphique 1) l'ensemble des films déjà vus, douze films particulièrement aimés sur les treize proposés (seul *Titanic* est à droite) et les fréquences de sorties au cinéma les plus élevées (entre cinq et onze fois par an ; au moins une fois par mois).

Graphique 1 – L'espace des goûts cinématographiques (axes 1-2 ; ACM « classique »)



Mais c'est avant tout le fait d'avoir vu ou non les films proposés qui est au principe de la première dimension de l'espace des goûts cinématographiques : les treize variables sur les films déjà vus contribuent pour 62 % à la construction du premier axe, les variables sur les films aimés particulièrement pour 12 %, et la variable sur la fréquence des sorties pour 4 %. Cette dimension d'intensité de la pratique est étroitement associée à l'âge, au sexe et au niveau de diplôme des enquêtés (Graphique 2) : les modalités

8 Pour une analyse approfondie de l'offre et des goûts cinématographiques en France, voir Duval, 2011.

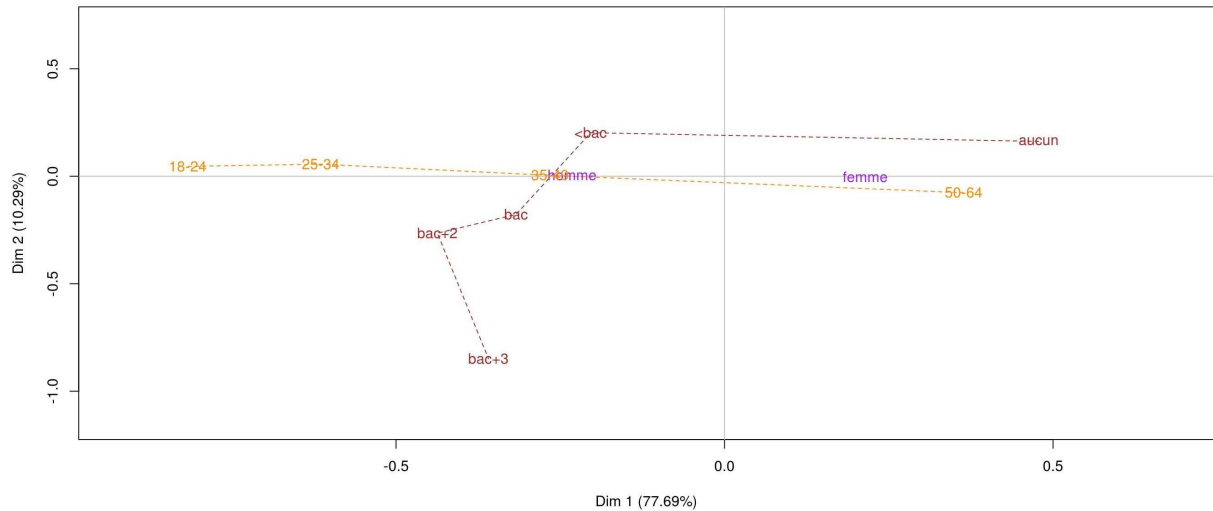
9 « que vous préférez regarder, que ce soit chez vous ou au cinéma »

10 « parce que vous savez qu'ils ne vous plaisent pas »

11 Pour une présentation détaillée de l'ACM et de l'analyse géométrique des données en général, voir Le Roux et Rouanet, 2004.

d'âge sont presque parfaitement alignées avec l'axe et « expliquent » 41 % de la variance entre les individus ; les hommes s'opposent nettement aux femmes (5% de la variance) et les diplômés aux sans-diplôme (14 % de la variance). Ce sont les femmes, les plus de 50 ans et les sans diplôme qui se situent du côté d'une faible intensité de la pratique cinématographique.

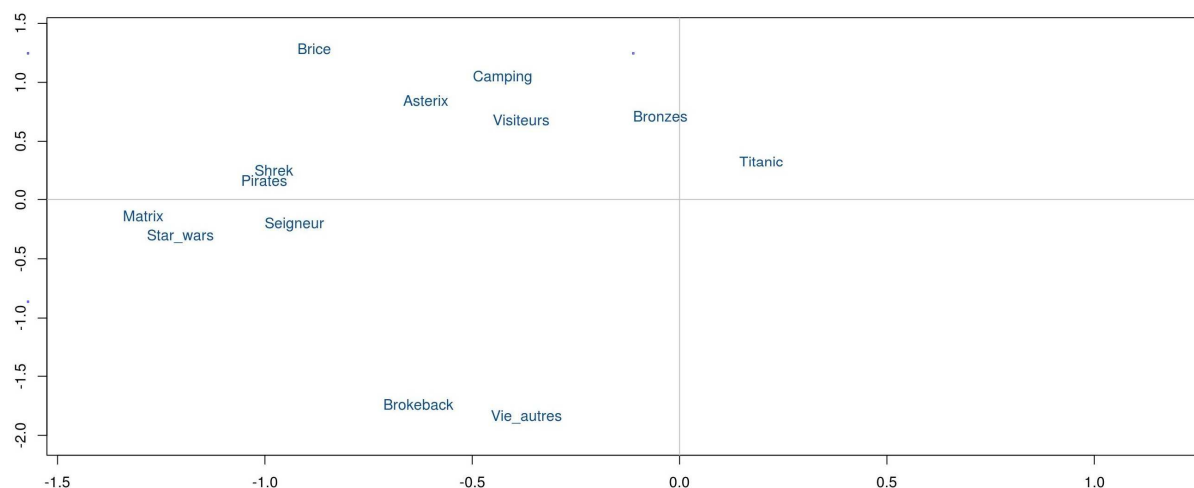
Graphique 2 – Les facteurs structurants de l'espace des goûts cinématographiques (axes 1-2 ; ACM « classique »)



Le second axe semble dessiner une opposition entre le goût pour les comédies françaises grand public (*Brice de Nice*, *Camping*, *Astérix et Obélix*, *Les Bronzés*, *Les Visiteurs*) et le goût pour des films d'auteur étrangers (*La vie des autres*, *Le secret de Brokeback Mountain*) : cette dimension échelonne les goûts en termes de *légitimité*, les films d'action ou de science-fiction étatsuniens se situant dans une position intermédiaire. On note d'ailleurs que les variables de « goûts » (films particulièrement aimés), exprimant plus directement la dimension distinctive de la légitimité culturelle, contribuent plus à la construction de ce second axe (39%) que les variables de pratique (films déjà vus ; 22%). Dans la même logique, les variables de « dégoût » pour les films font une apparition notable (15 %, contre 3 % pour le premier axe). Sans surprise, on observe une nette homologie entre la distribution du capital culturel (sous sa forme scolaire) et la légitimité des goûts : le niveau de diplôme explique 11 % de la variance du nuage des individus sur le second axe. En revanche, le poids de l'âge et du sexe comme facteurs structurants est ici négligeable.

On s'intéresse maintenant au goût pour les différents films proposés dans l'espace des goûts cinématographiques, et plus particulièrement dans la dimension de l'intensité de la pratique (Graphique 3).

Graphique 3 – Les films particulièrement aimés dans l'espace des goûts cinématographiques (axes 1-2 ; ACM « classique »)



Les films d'action et de science-fiction étatsuniens forment un groupe relativement homogène à l'extrême gauche du graphique, tandis que les comédies françaises (en haut) et les films d'auteur (en bas) sont plus en retrait. Cela semble indiquer que le goût pour les films d'action et de science-fiction est le fait des amateurs les plus assidus de cinéma, alors que les autres films sont plus appréciés par des individus aux pratiques moins intenses. Cependant, on sait par ailleurs (ministère de la Culture et de la Communication-DEPS, 2008) que ce sont les femmes de plus de trente ans qui ont les goûts les plus légitimes en matière de cinéma (comme dans d'autres domaines culturels). Or c'est aussi cette catégorie qui regarde le moins de films, comme on l'a montré plus haut. Il se pourrait donc que l'association observée entre le goût pour les films d'auteur et une pratique cinématographique relativement peu intense soit le résultat d'un effet de la structure d'âge et de sexe de la population des amateurs de cinéma légitime.

### Étudier les effets de structure : l'Analyse Factorielle Standardisée

Pour en avoir le cœur net, il faudrait pouvoir contrôler la structure de sexe et d'âge. Dans le cadre de l'analyse géométrique des données, cela impliquerait de contraindre les axes de l'ACM à être indépendants (i.e. orthogonaux) des caractéristiques de sexe et d'âge des individus, c'est-à-dire à construire une ACM « toute chose (d'âge et de sexe) égale par ailleurs »<sup>12</sup>. Pour ce faire, nous proposons une démarche en plusieurs étapes, que nous nommerons « Analyse Factorielle Standardisée » (AFS)<sup>13</sup>, par analogie avec les méthodes de standardisation en démographie (Léridon et Toulemon, 1997 ; Deauvieux, 2011).

- Dans un premier temps, on part de l'ACM « classique » réalisée précédemment pour construire un nouveau tableau de données (appelé C) réunissant les coordonnées des individus sur chacun des axes de l'ACM. On a donc un tableau Individus x Coordonnées, avec  $i$  lignes et  $p$  colonnes ( $i$  est le nombre d'individus et  $p$  le nombre de composantes principales de l'ACM)<sup>14</sup>.
- Ensuite, on réalise  $p$  régressions linéaires, i.e. une par colonne du tableau C. Les coordonnées individuelles de l'ACM (i.e. les variables de C) sont utilisées l'une après l'autre comme variable dépendante, avec le sexe et l'âge systématiquement en variables indépendantes.
- Pour chacune des régressions, on récupère les résidus : on a alors un nouveau tableau (appelé R) avec  $i$  lignes et  $p$  colonnes. Les variables quantitatives qui composent R correspondent aux coordonnées des individus dans l'ACM initiale « purgées » de la structure de sexe et d'âge.
- On réalise une Analyse en Composantes Principales (ACP) à partir du tableau R. On obtient ainsi un nouveau nuage des individus, i.e. un nouvel espace.
- Si l'on projette en variables supplémentaires sur cet espace les variables qui ont servi à construire l'ACM initiale, on obtient un espace des goûts cinématographiques « net » de la structure de sexe et d'âge des individus<sup>15</sup>.

Lorsque l'on applique une Analyse Factorielle Standardisée (par le sexe et l'âge) à nos données, le résultat est sans équivoque (Graphique 4). On a toujours trois groupes de films relativement homogènes, mais les

12 Pratiquement, cela signifie que toutes les modalités de sexe et d'âge projetées en variables supplémentaires apparaissent au centre du nuage des individus.

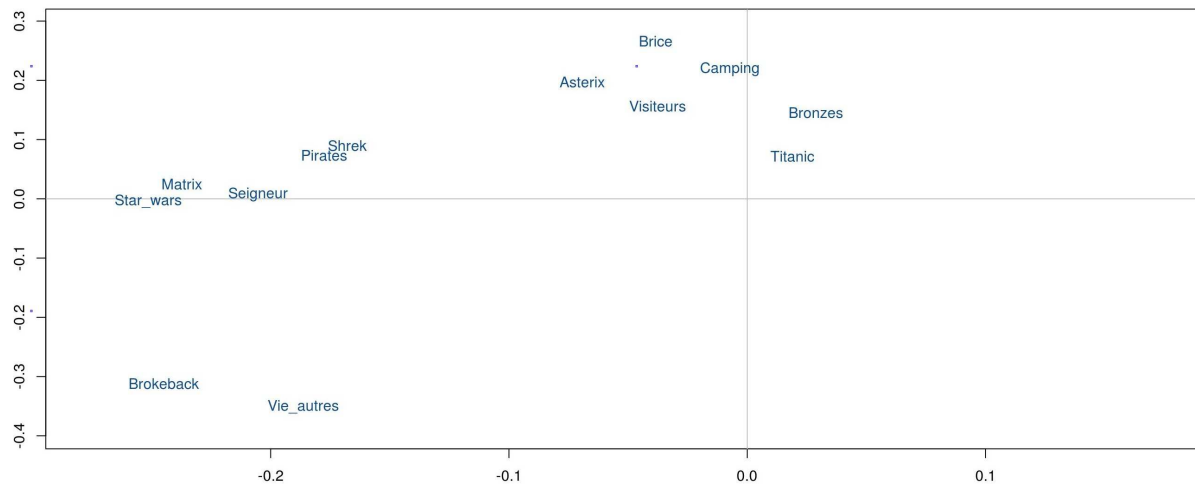
13 Cette méthode est implémentée dans le package R « GDAtools » en cours de développement et disponible sur la page web suivante : <http://nicolas.robette.free.fr/outils.html>

14 On peut lors de cette étape choisir de ne retenir que les premières composantes principales, celles qui contiennent le plus d'information.

15 Dans la mesure où l'ACP n'est pas réalisée directement sur les variables de l'ACM initiale, on ne peut calculer les contributions de ces variables. En revanche, toutes les aides à l'interprétation des variables supplémentaires sont utilisables ( $\cos^2$ , v-test,  $\eta^2$ ).

comédies françaises se sont légèrement déplacées vers la droite du premier axe, et surtout les deux films d'auteur se situent maintenant à l'extrême gauche du graphique, au même niveau d'intensité de la pratique cinématographique que les films d'action et de science-fiction<sup>16</sup>. La comparaison des deux espaces, l'un avec les effets de structure et l'autre sans, montre ainsi qu'à sexe et âge équivalents, les amateurs de films légitimes ont une pratique aussi intense que les amateurs de films d'action. Ou, pour le dire autrement, l'association observée entre le goût pour les films d'auteur et un niveau de pratique relativement limité est le résultat de la structure de sexe et d'âge de la population concernée.

Graphique 4 – Les films particulièrement aimés dans l'espace des goûts cinématographiques (axes 1-2 ; Analyse Factorielle Standardisée)



## Conclusion

Lorsque l'on souhaite analyser des effets de structure dans un cadre d'analyse géométrique des données, une première ligne de recherche consiste à visualiser graphiquement les résultats d'une régression (Rouanet *et al*, 2002 ; Lebaron, 2013). Il s'agit de construire l'espace multidimensionnel formé par les variables indépendantes du modèle et de projeter ensuite la variable dépendante dans cet espace : on peut alors comparer les effets globaux et partiels des variables indépendantes. Nous proposons une ligne de recherche complémentaire, l'AFS, qui part cette fois d'une analyse géométrique – ici une ACM<sup>17</sup> – et utilise dans un second temps la régression pour mettre au jour des effets de structure dans l'espace initialement construit. Le principal point commun des deux approches est qu'elles visent à comparer les résultats avec et sans effet de structure et non à éliminer ceux-ci, et ce sans modèle aléatoire ni cadre probabiliste *a priori* (Lebaron, 2013).

L'Analyse Factorielle Standardisée incite par ailleurs à une réflexion plus générale. Les nombreuses applications que nous en avons tentées (et dont nous ne pouvons rendre compte dans le détail dans les limites de cet article) mettent toutes en évidence le même phénomène : lorsque l'on neutralise des effets de structure, l'espace multidimensionnel initial, dans ses principes généraux, reste pratiquement identique. Certaines modalités peuvent se déplacer, comme les films d'auteur dans l'exemple présenté ici, mais la structure globale du nuage des modalités et l'interprétation que l'on peut donner de ses différentes

16 Les autres modalités du goût cinématographiques ne sont pas représentées ici pour faciliter la lecture mais leur structure globale est la même qu'avec l'ACM « classique », et l'interprétation des deux premiers axes (intensité ; légitimité) est inchangée.

17 Dans la mesure où dans notre approche les régressions sont réalisées à partir des coordonnées des individus sur les axes, la méthode proposée peut également s'appliquer à d'autres techniques d'analyse géométrique : Analyse en Composantes Principales (ACP), Analyse Factorielle Multiple (AFM), Multidimensional Scaling (MDS), etc.

dimensions « résistant », y compris lorsque l'on contrôle conjointement par plusieurs variables fortement « explicatives » (telles que la classe sociale, le genre, l'âge et le capital scolaire). Cela illustre sans doute une hypothèse épistémologique plus générale : les déterminismes sociaux, même solidement établis, sont probabilistes, et non mécaniques, et partiels, au sens où ils ne produisent leurs effets qu'associés les uns aux autres. Les corrélations observées empiriquement entre des propriétés sociales, formant par exemple des univers de goûts relativement cohérents, peuvent être expliquées par divers déterminants sociaux, mais ces déterminants n'« épuisent » jamais totalement la réalité sociale. Les effets de structure massifs, avec par exemple des retournements comme c'est le cas de la réussite scolaire chez les enfants d'immigrés dans l'article de Vallet et Caille (1995), sont finalement assez rares. Ce qui ne réduit en rien la nécessité d'analyser ces effets de structure !



## Bibliographie

- Abbott A. 2001. *Time matters. On theory and method*. Chicago: The University of Chicago Press.
- Bourdieu P., 2000. *Les structures sociales de l'économie*. Paris : Seuil (coll. Liber).
- Bourdieu P., Darbel A., 1966. « La fin d'un malthusianisme ». In Darras, *Le partage des bénéfices*. Les Éditions de Minuit, p. 135-154.
- Deauvieu J., 2011. « Est-il possible et souhaitable de traduire sous forme de probabilités un coefficient logit ? Réponse aux remarques formulées par Marion Selz à propos de mon article paru dans le BMS en 2010 ». *Bulletin de Méthodologie Sociologique*, 112 (1): 32-42.
- Desrosières A., 2003. « Bourdieu et les statisticiens. Une rencontre improbable et ses deux héritages ». In P. Encrevé R.M. Lagrave, *Travailler avec Bourdieu*. Paris: Flammarion.
- Desrosières A., 2001. « Entre réalisme métrologique et conventions d'équivalence : les ambiguïtés de la sociologie quantitative ». *Genèses*, (43): 112-127.
- Donnat O., 2009. *Les pratiques culturelles des Français à l'ère numérique. Enquête 2008*. Paris : La Découverte.
- Duflot E., 2010. *Le Développement humain. Lutter contre la pauvreté (I)*. Paris : Le Seuil / République des idées.
- Duval J., 2011. « L'offre et les goûts cinématographiques en France ». *Sociologie*, 2(1), 1-18.
- Halbwachs M., 2002[1930]. *Les causes du suicide*. Paris : PUF.
- Le Roux B., Rouanet H., 2004. *Geometric Data Analysis: From Correspondence Analysis to Structured Data Analysis*. Springer.
- Lebaron F., 2010. « L'analyse géométrique des données dans un programme de recherche sociologique : Le cas de la sociologie de Bourdieu ». *Modulad*, 42.
- Lebaron F., 2013. « La régression peut-elle faire progresser ? Approches et usages critiques des modèles de régression ». Présentation à l'atelier « Sociologie quantitative et sociologie de la quantification », laboratoire Printemps, 22 mars.
- Lebart L., Morineau A., Piron M., 2000. *Statistique exploratoire multidimensionnelle*. Paris: Dunod.
- Léridon H., Toulemon L., 1997. *Démographie. Approche statistique et dynamique des populations*. Paris : Economica.
- Lieberson S., 1995. *Making It Count. The Improvement of Social Research and Theory*. University of California Press.
- Ministère de la Culture et de la Communication – DEPS, 2008. *Les résultats complets de l'enquête 2008*. <http://www.pratiquesculturelles.culture.gouv.fr/08resultat.php>
- Nétumières (des) F., 1997. « Méthodes de régression et analyse factorielle ». *Histoire & Mesure*, 12(3-4), 272-297.
- Ollion E., 2011. « De la sociologie en Amérique. Éléments pour une sociologie de la sociologie étasunienne contemporaine ». *Sociologie*, 2(3) : 277-294.
- Passeron J.-C., 1991, *Le raisonnement sociologique. L'espace non-poppérien du raisonnement naturel*, Paris, Nathan.
- Pénissat E., 2011. « Quantifier l'effet "pur" de l'action publique : entre luttes scientifiques et redéfinition des politiques d'emploi en France ». *Sociologie et sociétés*, 43(2), 223-247.
- Ragin C., 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. University of California Press.

Rouanet H., Ackermann W., Le Roux B., 2000. « The Geometric Analysis of Questionnaires: The Lesson of Bourdieu's La Distinction ». *Bulletin de Méthodologie Sociologique*, 65 (1): 5-18.

Rouanet H., Lebaron F., 2006. « La preuve statistique: regard critique sur la régression ». Présentation au séminaire "Qu'est-ce que Faire preuve?", Amiens.

Rouanet H., Lebaron F., Hay V.L. , Ackermann W., Le Roux B., 2002. « Régression et analyse géométrique des données: réflexions et suggestions ». *Mathématiques et sciences humaines*, (160): 13-46.

Saint-Martin (de) M., Bourdieu P., 1976. « Anatomie du goût ». *Actes de la recherche en sciences sociales*, 2(5), 2-81.

Vallet L.-A., 2004. « A propos d'un ouvrage peu connu dans la sociologie française:" Making It Count. The Improvement of Social Research and Theory" de Stanley Lieberson ». *Revue européenne des sciences sociales*, XLII (129): 341-348.

Vallet L.-A., Caille J.-P., 1995. « Les carrières scolaires au collège des élèves étrangers ou issus de l'immigration ». *Éducation et Formations*, (40): 5-14.