

Démographie et différences

Colloque international de Montréal (7-10 juin 1988)



ASSOCIATION INTERNATIONALE DES DÉMOGRAPHES DE LANGUE FRANÇAISE

AIDELF

Problèmes d'interprétation inhérents à l'usage de comparaisons

- Jacques MENTHONNEX
SCRIS, Lausanne, Suisse

Intérêt offert par l'étude des contrastes

Les recherches effectuées en vue d'améliorer notre connaissance concernant certains phénomènes socio-démographiques conduisent à utiliser deux voies qui sont complémentaires. L'une est celle de la démarche expérimentale basée, en sciences sociales, sur l'observation; l'autre est celle de la conceptualisation au moyen d'une démarche théorique qui s'appuie sur des règles de logique et qui peut être formalisée, par exemple, au moyen d'un modèle mathématique qui inclut un ensemble d'hypothèses et de propositions.

La démarche théorique est élaborée à partir d'une certaine connaissance a priori. Elle offre la possibilité d'une confrontation, souhaitable, entre le comportement du modèle et notre perception de la réalité; si l'observation n'infirme pas la théorie sous-jacente au modèle, notre connaissance pourra alors être enrichie.

Pratiquement, les démarches expérimentales et théoriques sont parfois confondues lorsqu'il y a conceptualisation d'un phénomène aisément descriptible et proche de l'intuition.

La démarche expérimentale est importante car c'est elle qui guide le choix d'hypothèses nécessaires pour une conceptualisation du phénomène étudié. Cette démarche devrait suivre un plan d'expérience. En fait, le plus souvent seule intervient l'intuition du chercheur, guidé parfois par les outils de l'analyse multivariée.

La démarche expérimentale nécessite de s'assurer préalablement de la cohérence de l'expérience (problème de mesure, comparabilité des concepts, etc.) et recourt fréquemment à une analyse comparative.

- La comparaison est nécessaire pour conférer aux faits une intelligibilité qu'ils n'ont pas naturellement. Comme il est difficile de retirer d'une image unique une perception des mécanismes vraisemblables, on a souvent recours à une comparaison dans le temps ou dans l'espace.

- En s'assurant de certaines analogies entre les comportements de plusieurs sous-groupes, on peut constater qu'une liaison n'est pas seulement fortuite.

- La comparaison peut éliminer certaines idées préconçues et peut ouvrir des pistes de recherche. La juxtaposition de trajectoires particulières facilite la découverte de phénomènes généraux.

Difficultés d'agréations inhérentes à l'existence de contrastes

La représentation d'idées ou de connaissances relatives à un phénomène concernant la collectivité est parfois formulée au moyen d'un modèle qui explicite la liaison entre des variables endogènes ou *expliquées* (Y) et des variables exogènes ou *explicatives* (X) :

$$Y = F(X) + e$$

La liaison est supposée définie aux écarts aléatoires près (e), qui traduisent le fait que le modèle ne décrit pas la réalité parfaitement et exhaustivement. En général, F correspond à un modèle agrégé (macro) dans la mesure où il reflète (ou devrait refléter) le comportement d'une collectivité. Il faut remarquer que le choix des variables exogènes et le choix de la forme analytique de F dépendent fréquemment de l'intuition du chercheur, intuition qui se raccroche le plus souvent à un environnement très désagrégé (micro), car un individu présentera plus spontanément un phénomène au niveau individuel qu'au niveau collectif.

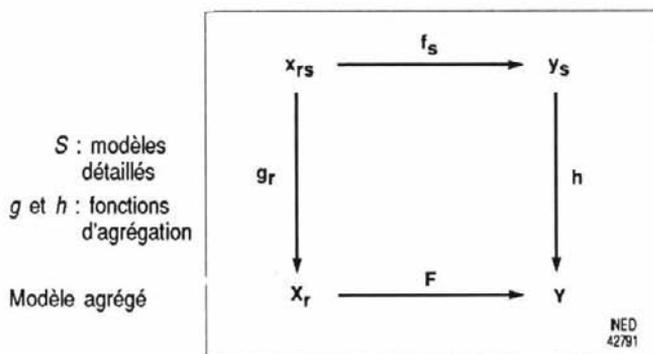
Plusieurs préoccupations liées à l'existence possible de contrastes à l'intérieur de la collectivité considérée peuvent alors apparaître.

Il est souvent difficile d'avoir une idée très précise sur la forme de la fonction analytique F et l'on peut déjà être satisfait si l'on arrive à mettre en évidence l'existence de corrélations non fortuites entre l'évolution de plusieurs variables. On choisit alors une fonction simple et linéaire (ou linéaire à une transformation biunivoque près) permettant d'étudier les caractéristiques des paramètres après avoir supposé que les écarts e suivent une loi de probabilité normale d'espérance mathématique nulle.

Cette façon de faire sous-entend le plus souvent que l'on considère que le comportement étudié est homogène à l'intérieur du groupe considéré. Si, pour le sous-groupe s ($s = 1, \dots, S$), la variable endogène y_s est dépendante des variables x_{rs} ($r = 1, \dots, R$), on suppose implicitement que

$$Y = \Sigma Y_s \quad \text{et que} \quad \frac{\partial Y}{\partial X_r} \equiv \frac{\partial y_s}{\partial x_{rs}}$$

c'est-à-dire, pour un modèle linéaire, que les paramètres estimés au niveau agrégé sont un bon reflet de la liaison entre y_s et chacune des variables explicatives de chaque modèle détaillé ($s = 1, 2, \dots, S$).



En fait, si l'on veut proposer une interprétation correcte des paramètres estimés au niveau agrégé, il est nécessaire de savoir s'il existe des contrastes de comportement importants à l'intérieur du groupe étudié. Sur ce point, l'étude des problèmes d'agrégation est riche d'enseignements. Le schéma ci-contre permet d'illustrer le contexte.

Un des problèmes consiste à chercher dans quelles conditions il existe une relation F entre les variables endogènes agrégées Y et les variables exogènes agrégées X qui soit compatible avec les modèles détaillés.

Un cas particulier permet de montrer que ces conditions sont très restrictives.

Si l'on choisit les fonctions d'agrégation g et h telles que :

$$Y = \sum_s \alpha_s y_s \quad \text{et} \quad X_r = \sum_s \omega_{rs} x_{rs} \quad r = 1, \dots, R$$

les conditions nécessaires et suffisantes d'agrégation parfaite sont les suivantes (application du théorème de A. Nataf) :

$$\frac{\partial F}{\partial X_r} = \frac{\alpha_s}{\omega_{rs}} \cdot \frac{\partial f_s}{\partial x_{rs}} = \frac{\alpha_s}{\omega_{rs}} \cdot \frac{\partial f_s}{\partial x_{rs}} \quad \forall t \neq s = 1, \dots, S$$

Cela impose une liaison entre les dérivées $\frac{\partial f_s}{\partial x_{rs}}$ de chaque sous-groupe ; on obtient :

$$y_s = a_s + \sum_r c_r \cdot \frac{\omega_{rs}}{\alpha_s} \cdot x_{rs}$$

$$\text{et } Y = a + \sum_r c_r \cdot X_r \quad \text{où } a = \sum_s a_s \cdot \alpha_s$$

Or les coefficients de pondération ω_{rs} et α_s sont en général donnés a priori, suivant la signification des variables. Cela sous-entend que lorsqu'il existe des contrastes de comportement entre les S sous-groupes, il est pratiquement impossible de trouver un modèle agrégé qui soit cohérent avec l'existence de ces contrastes.

Ce constat amène naturellement quelques réflexions concernant la signification des modèles macro :

- Lorsque la spécification d'un modèle est choisie à partir de considérations théoriques, il est nécessaire de tester et d'utiliser le modèle au même niveau d'agrégation que celui prévu dans la théorie sous-jacente.

- Lorsque la spécification d'un modèle est choisie empiriquement, la signification du modèle spécifié peut varier de cas en cas :

- a) Si les comportements individuels (ou des sous-groupes) sont homogènes, une liaison statistiquement établie au niveau global entre variables exogènes et variables endogènes peut être le reflet d'une liaison effective de cause à effet entre les indicateurs étudiés. Le modèle contribue alors à conceptualiser une théorie relative à un phénomène.

- b) S'il existe des contrastes de comportements à l'intérieur du groupe retenu pour le modèle, de deux choses l'une :

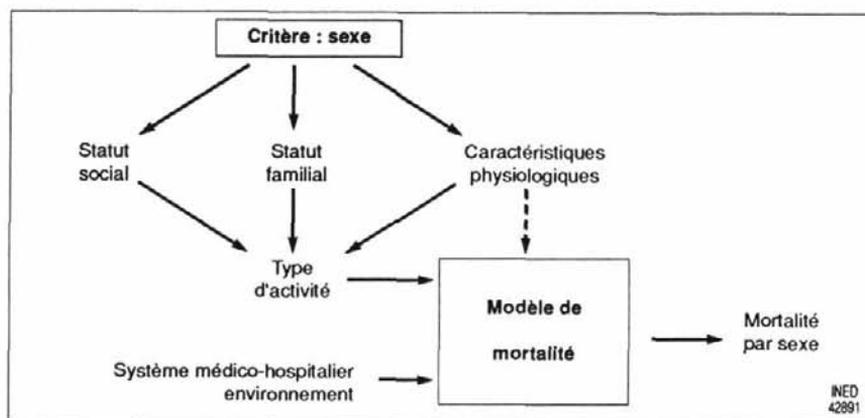
- ou bien on a de la chance, et les lois globales (macro) réalisent une synthèse des comportements individuels ; elles ont alors une nature probabiliste qui ne s'attache pas à exposer les mécanismes des actions individuelles (ou des sous-groupes homogènes),

- ou bien les lois globales n'expliquent rien, mais elles décrivent le fait que, dans un contexte donné, il y a eu telle ou telle coïncidence au niveau des indicateurs choisis.

Quelques préoccupations liées à l'interprétation des différences de comportement entre sous-groupes

L'analyse d'un phénomène à partir d'observations comporte le danger d'une interprétation trop hâtive, influencée par des intuitions mal fondées. Pour pallier en partie ce risque, il peut être avantageux de se rattacher à un schéma de base construit dans un contexte large (sous-entendu, qui nécessite le moins d'informations a priori possible). Cela permet alors de constituer, préalablement à l'analyse, un plan d'expérience. Même si cette préoccupation n'est pas nouvelle⁽¹⁾, de nombreux progrès restent à faire dans cette voie.

Dans le cadre de l'analyse comparative, un schéma simple peut déjà illustrer ce propos et rappeler qu'un contraste observé peut parfois ne dépendre qu'indirectement du critère qui a délimité les sous-groupes étudiés :



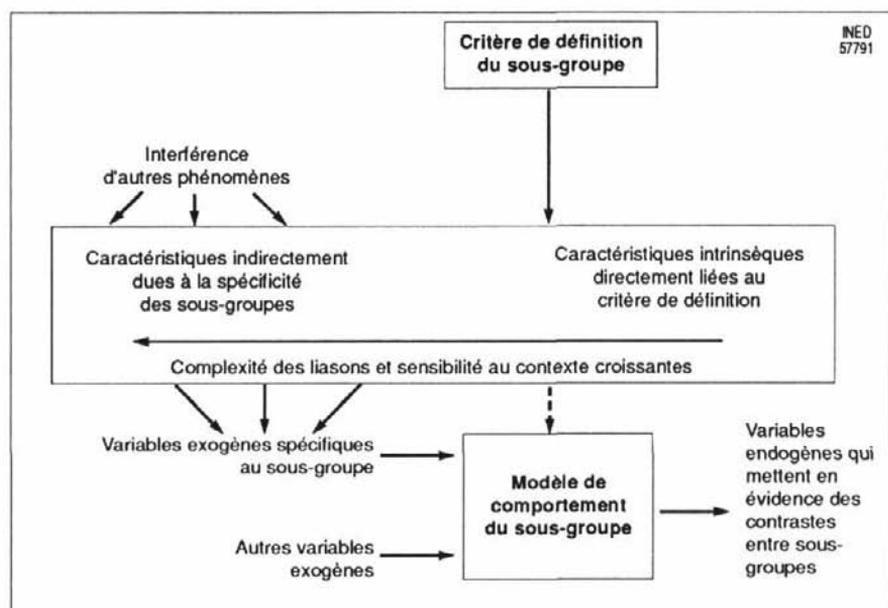
Dans le contexte de l'analyse de la mortalité différentielle par sexe, le critère de définition du sous-groupe (le sexe) peut avoir une incidence directe sur la mortalité, compte tenu des caractéristiques physiologiques propres à chacun des groupes. Par ailleurs, il peut exister des effets sur le niveau de la mortalité dus à des phénomènes moins immédiats, et spécifiques au contexte étudié comme, par exemple, l'activité qui est dépendante du sexe d'une manière spécifique à la région et à l'époque.

En généralisant, nous obtenons le schéma page suivante (qui ne met pas en évidence les interférences possibles entre les modèles des sous-groupes).

Ce schéma entraîne au moins trois remarques :

- Une première difficulté d'interprétation des différences de comportement est due au fait que les caractéristiques directement liées au critère de sélection choisi (par exemple, des caractéristiques physiologiques pour un critère « sexe », le climat pour un critère

⁽¹⁾ Voir John-Stuart Mill, chap. VIII de son *Système de logique* intitulé « Des quatre méthodes de recherche expérimentale ».



géographique, le salaire ou les risques encourus pour un critère professionnel) peuvent avoir une influence, tant sur certaines variables exogènes au modèle que sur la valeur des paramètres du modèle.

- Une deuxième difficulté est due à l'existence possible de phénomènes propres aux sous-groupes qui pourraient influencer certaines variables exogènes d'une manière inattendue par rapport au critère de sélection des sous-groupes. Par exemple, un critère géographique pourrait impliquer des niveaux particuliers de la proportion de femmes et créer ainsi des interférences sur le phénomène étudié; cela provoquerait alors des différences de comportement non généralisables. A l'extrême, il existe des interférences tellement indirectes qu'elles engendrent des écarts entre sous-groupes que l'on peut considérer comme aléatoires.

- Un autre problème peut encore apparaître dans la mesure où seules certaines associations bien précises entre les variables exogènes au modèle peuvent expliquer le comportement des sous-groupes. Les choix d'un modèle simplement additif risquent de faire aboutir à des conclusions sensiblement différentes de celles correspondant à un modèle multiplicatif par exemple. Comme, en général, un modèle additif est plus naturellement choisi (déjà au niveau intuitif) à cause de ses propriétés de séparabilité, on peut se demander dans quelle mesure des réflexes de simplification ne nous induisent pas en erreur en occultant certains effets synergiques importants.

BIBLIOGRAPHIE

- BERTHOUD G. et BUSINO G. (édité par) «La comparaisons en sciences humaines et sociales» Colloque 1985 de Lausanne, dans Cahier Vilfredo Pareto 1986 N 72 – libr. Droz – Genève.
- BOUDON R. et LAZARFELD P. (texte choisis par) *L'analyse empirique de la causalité*, 1966 – Mouton & Co – Paris.
- GREEN H.A. *Aggregation in Economic Analysis, An Introductory Survey*, Princetown University Press – 1964.
- GUPTA K.L. *Aggregation In Economics. A theoretical and empirical study*, 1969 –Rotterdam University Press.
- MILL John-Stuart *Système de logique déductive et inductive – exposé des principes de la preuve et des méthodes de recherche scientifique*, 1843. Traduction L. Peisse 1866 – Libr. Philosophique de Ladrage – Paris – 2 vol.
- SOLARI Luigi *De l'économie qualitative à l'économie quantitative*, (chap. 9 : «Problèmes d'agrégation») 1977 – Masson –Paris.
- SOLARI Luigi *Essais de méthode et analyses économétriques*, 1979 – Droz – Genève.