



Séminaire de Méthodologie Statistique

Mardi 5 juillet 2016

14h-17h, Insee - Malakoff 1 - salle Malinvaud (1245)

Géographie et statistique :
en amont et en aval, des questions qui ne
coulent pas de source

La question de la mobilisation de l'information géographique pour améliorer le processus de production statistique fait l'objet de très nombreuses initiatives internationales et nationales. En amont, il s'agit de mettre en place une organisation assurant une meilleure communication entre les systèmes d'information géographique et statistique au travers d'un renforcement des liens entre instituts de géographie et de statistique, d'assurer une standardisation et une interopérabilité des données, de mettre en place des algorithmes de géoréférencement. En aval, il s'agit de mobiliser l'information géographique pour alimenter la demande croissante d'information statistique finement localisée. L'Insee est impliqué dans ces réflexions, que ce soit à l'ONU, à Eurostat, et/ou en liaison avec l'IGN.

Après une introduction présentant quelques enjeux actuels autour de la géographie dans le domaine de la statistique publique, le séminaire se propose d'illustrer le processus de géo-référencement par l'exemple d'une source administrative, FiLoSoFi. Dans une deuxième partie, on présentera la spécificité des données géolocalisées, à savoir la dépendance spatiale, et la mesure de l'intensité de cette dépendance. Dans une troisième partie, les moyens pour prendre en compte cette dépendance dans la méthodologie statistique seront exposés dans le cadre de l'économétrie.

Méthodologie de géolocalisation des ménages fiscaux dans FiLoSoFi

Ali Hachid - *Division Méthodes et référentiels géographiques, Insee*

Introduction à l'analyse spatiale

Marie-Pierre de Bellefon - *Division Méthodes et référentiels géographiques, Insee*

Économétrie spatiale : une introduction

Ronan Le Saout - *Division Méthodes appliquées de l'économétrie et de l'évaluation, Insee*

Jean-Michel Floch - *Département de l'action régionale, Insee*

Géographie et statistique : en amont et en aval, des questions qui ne coulent pas de source

Résumés des interventions

Méthodologie de géolocalisation des ménages fiscaux dans FiLoSoFi

Ali Hachid - *Division Méthodes et référentiels géographiques, Insee*

Le dispositif FiLoSoFi (Fichier Localisé Social et Fiscal), qui se substitue aux dispositifs RFL (Revenus Fiscaux Localisés) et RDL (Revenus Disponibles Localisés) à partir du millésime 2012, vise à mesurer le revenu disponible, le taux de pauvreté et les niveaux de vie des ménages à partir de l'exploitation des diverses sources fiscales et sociales, et couvre tous les niveaux géographiques infra-nationaux, de la région au quartier de la politique de la ville ou à l'Iris, en passant par la commune. Afin de pouvoir décliner les indicateurs de distribution de niveaux de vie à des échelons géographiques fins, en particulier les Iris et les nouveaux quartiers de la politique de la ville (QPV), des travaux ont été menés pour attribuer un code Iris et un code QPV aux ménages en exploitant deux informations géographiques présentes dans la base : l'adresse et la parcelle cadastrale.

La parcelle cadastrale n'étant pas exploitée dans les anciens dispositifs pour la diffusion infracommunale, la méthodologie mise en place pour FiLoSoFi permet alors de réduire très significativement la charge de reprise manuelle. Parallèlement, elle permet de montrer les limites de l'utilisation des parcelles cadastrales pour repérer finement les ménages. Enfin, l'introduction d'une diffusion sur les nouveaux périmètres de la politique de la ville conjointement à la diffusion de niveau Iris nécessite de contrôler le risque de rupture de confidentialité par différenciation.

Introduction à l'analyse spatiale

Marie-Pierre de Bellefon - *Division Méthodes et référentiels géographiques, Insee*

L'analyse spatiale s'intéresse au lieu où les choses se produisent et aux relations entre elles. Le lieu est référencé dans un système d'information géographique grâce à ses coordonnées. Une des caractéristiques de l'analyse spatiale est donc que le support de l'observation contient des informations.

Les données géolocalisées peuvent être réparties en deux catégories : données surfaciques et données ponctuelles. La différence fondamentale entre ces données n'est pas la taille de l'unité géographique considérée mais le processus générateur des données. Pour des données surfaciques, la localisation des observations est considérée comme fixe : c'est la valeur des observations qui suit un processus aléatoire. Au contraire, pour des données ponctuelles, c'est la localisation des observations qui est aléatoire. Cette présentation s'intéresse aux méthodes d'analyse des données surfaciques.

Pour étudier les relations spatiales entre les objets, il faut d'abord définir la structure spatiale dans laquelle ils s'insèrent : choix de l'échelle d'agrégation et du type de zonage, puis définition du voisinage d'une observation (en fonction de la distance, de la contiguïté, ou d'autres caractéristiques de données).

Dans un deuxième temps, les indices d'autocorrélation spatiale permettent de mesurer la dépendance spatiale entre les observations. La dépendance spatiale désigne une situation où la valeur d'une observation est liée aux valeurs des observations voisines (soit elles s'influencent mutuellement, soit elles sont toutes les deux soumises à un même phénomène non observé). Cette présentation introduit les principaux indices d'autocorrélation spatiale : indices de Moran, de Geary, de Getis et Ord. On étudie leurs propriétés, leur interprétation à une échelle globale ou locale, ainsi que l'influence du choix de la structure spatiale sur les résultats.

Géographie et statistique : en amont et en aval, des questions qui ne coulent pas de source

Insee - Département des méthodes statistiques

Séminaire du 5 juillet 2016

Économétrie spatiale : une introduction

Ronan Le Saout - *Division Méthodes appliquées de l'économétrie et de l'évaluation, Insee*

Jean-Michel Floch - *Département de l'action régionale, Insee*

Cette communication décrit la conduite d'une étude d'économétrie spatiale, à travers une modélisation descriptive du taux de chômage par zone d'emploi. Les modèles spatiaux ont néanmoins une application plus large, l'approche étant compatible avec tout problème où des relations de « voisinage » interviennent. La théorie économique caractérise en effet de nombreux cas d'interactions entre agents (produits, entreprises, individus), qui ne sont pas de nature géographique.

La communication se concentre sur l'étude de la dépendance spatiale, et donc sur ces interactions, et détaille les liens avec l'hétérogénéité spatiale, à savoir les phénomènes différenciés spatialement. Plusieurs formes d'interactions existent, relatives à la variable à expliquer, aux variables explicatives ou aux variables inobservées. De nombreux modèles se retrouvent donc en concurrence, à partir d'une même définition préalable des relations de voisinage. Une méthodologie pas à pas de choix de modèle (estimation et tests) est détaillée. Des effets de rétroaction entraînent une interprétation particulière (et plus complexe) des résultats.