

PRÉSENTATION

Olivier Sautory
Chef de la Division "Méthodologie d'élaboration
et d'analyse des données" de l'Insee

Les cinquièmes "Journées de méthodologie statistique", organisées par l'Unité "Méthodes statistiques" de l'Insee, se sont déroulées dans les locaux de l'ENSAE les 11 et 12 décembre 1996. Le succès de cette manifestation est toujours grandissant, puisque près de 600 personnes se sont inscrites, la plupart d'entre elles ayant assisté à plusieurs sessions.

Cette année, une attention particulière a été portée aux questions de méthodologie de collecte de données, avec notamment des séances consacrées aux questionnaires, aux enquêtes sur des sujets sensibles, et à l'expérience canadienne d'enquêtes par téléphone, présentée par un collègue de Statistique Canada. D'autres thèmes furent abordés au cours de ces journées : des développements récents dans le domaine des séries temporelles, les derniers travaux de l'Insee sur le thème des statistiques locales, l'usage des indicateurs d'inégalité des distributions de revenus. La conférence inaugurale a traité des rapports étroits et complexes entre statistique et sociologie, et un second collègue canadien a abordé la question de la confidentialité des données à l'aide des méthodes de "brouillage".

Dans la conférence inaugurale, **Alain Desrosières** (Insee-Crest) a cherché à montrer la réciprocité des apports respectifs de la méthodologie statistique et de la sociologie. Dans la phase de construction des données, ces échanges peuvent porter sur la définition, la consistance et la fiabilité des variables, sur la structure et la signification des nomenclatures, sur les questions d'interaction entre enquêteurs et enquêtés, sur les difficultés du codage, ou même sur l'organisation du système statistique, administratif, universitaire ou privé. Dans la phase d'analyse des données, la coexistence ou la concurrence entre grandes familles de méthodologies peuvent elles-mêmes faire l'objet d'une analyse sociologique.

Conférences spéciales

Dans la première conférence spéciale, "Confidentialité des données ou l'art du brouillage clair", **Jean-René Boudreau** (Statistique Canada) a exposé comment son institut aborde le problème de la confidentialité des réponses dans les produits de diffusion appelés "fichiers de microdonnées". Pour assurer cette confidentialité, on peut être amené à modifier les réponses (c'est-à-dire introduire du bruit dans les données) de certains répondants pour réduire ou éliminer la possibilité d'identification nécessaire à toute divulgation. Cependant, ces modifications ne doivent pas invalider l'utilisation des données. D'où les questions : comment peut-on brouiller les données d'une manière "intelligible" ? Doit-on brouiller les données globalement (toutes les données) ou localement (seulement certaines données) ? Comment choisit-on les données qui seront brouillées ?

La pratique des enquêtes téléphoniques a, au cours des années, pris de plus en plus d'importance au sein du programme d'enquêtes de Statistique Canada. Plusieurs raisons peuvent expliquer ce fait. Cette méthode offre à la fois flexibilité, efficacité, et rapidité d'exécution, et peut, dans plusieurs situations, se substituer avantageusement aux interviews sur place très coûteuses. Par ailleurs, elle constitue un complément essentiel aux enquêtes postales, et se prête très bien aux méthodes assistées par ordinateur. **Jean-François Gosselin** (Statistique Canada) a présenté une vue d'ensemble de la pratique des enquêtes téléphoniques dans son institut, et a indiqué les questions et inquiétudes qui influenceront grandement sur l'avenir de ce mode de collecte, concernant en particulier le taux de réponse aux enquêtes.

Les questionnaires et réponses aux enquêtes

La qualité d'une enquête est liée en particulier à la qualité du questionnaire utilisé pour la collecte. La conception et l'évaluation du questionnaire sont donc des étapes très importantes dans le développement d'une enquête. **France Bilocq** (Direction Générale de l'Insee (DG), unité méthodes statistiques, et Statistique Canada) a mis en évidence un certain nombre de principes de base destinés à minimiser les erreurs de réponse, en axant son intervention sur la formulation des questions et sur les différentes méthodes d'évaluation des questionnaires. La nécessité de la mise à profit de l'expérience des enquêteurs lors de cette phase de l'enquête a été soulignée.

Dans les enquêtes par sondage auprès des ménages ou des individus, de nombreuses questions peuvent être posées mais sur un échantillon de taille réduite. Au contraire, le caractère exhaustif d'un recensement de la population, s'il autorise des analyses sur des sous-populations très fines, limite l'ampleur et la complexité du questionnement. **Jacqueline Lacroix** (DG, division recensements de la population) a fait le point sur l'élaboration des questionnaires du 33^{ème} recensement de la

population. Elle a présenté en particulier les résultats de tests de collecte permettant de juger de l'acceptabilité et de la compréhension de nouvelles questions, retenues après une large consultation, menée dans le cadre du CNIS auprès des principaux utilisateurs des données.

Depuis 1993, plusieurs enquêtes auprès des ménages réalisées par l'Insee se sont vues refuser par la CNIL le statut obligatoire, jusqu'alors systématiquement accordé. Afin de mesurer l'impact de l'obligation de réponse sur le taux de réponse à ses enquêtes ménages, l'Insee a greffé un dispositif d'observation sur l'enquête "permanente conditions de vie des ménages" de janvier 1996 : deux échantillons de 2000 ménages ont été enquêtés dans des conditions très proches, la seule différence tenant à l'annonce du statut de l'enquête. **Catherine Berthier** et **Françoise Dupont** (DG, unité méthodes statistiques) ont présenté les premiers résultats de cette opération, dont le principal constat est le suivant : le comportement de refus des enquêtés au premier contact est nettement plus fréquent en l'absence de l'obligation (18,2% contre 8,6%), et la tentative de relance aboutit moins souvent.

Les séries temporelles

Deux nouveaux logiciels de désaisonnalisation sont apparus récemment : X12-ARIMA et TRAMO-SEATS. **Ketty Attal** (DG, unité méthodes statistiques) a consacré la première partie de son exposé à une présentation générale des méthodes de désaisonnalisation, du point de vue historique et du point de vue de leur philosophie. Elle a ensuite décrit les deux nouveaux logiciels, qui reposent sur des principes très différents, mais qui ont en commun d'une part l'utilisation des modèles ARIMA, d'autre part le souci d'améliorer la "qualité" de la série qui doit être décomposée, par un traitement préalable assez poussé.

Christophe Planas (Eurostat) s'est ensuite attaché à la comparaison de ces deux logiciels d'un point de vue théorique. Il a présenté les propriétés des différents filtres intervenant dans les logiciels, et les conséquences pratiques ont pu être vérifiées sur la série de la production industrielle mensuelle en France.

Les enquêtes de conjoncture réalisées par l'Insee offrent un type d'information particulièrement utile pour l'analyse conjoncturelle. Cependant, le nombre et la diversité des questions posées rendent souvent délicate l'interprétation des résultats obtenus. La communication de **Catherine Doz** (Direction de la Prévision) et **Fabrice Lenglard** (DG, division synthèse conjoncturelle) a présenté la construction d'un indice composite susceptible de résumer cette information, qui utilise les modèles à facteurs dynamiques. Les méthodes standard d'analyse factorielle classique ont également été utilisées, conduisant à des résultats très proches des précédents, sur les données des enquêtes mensuelle et trimestrielle de conjoncture dans l'industrie.

Les mesures d'inégalité

Dans un exposé introductif, **Olivier Sautory** (DG, unité méthodes statistiques) a présenté les principaux indicateurs utilisés pour mesurer l'inégalité (dans le cas de distributions de revenus en particulier), ainsi que leur interprétation et leurs propriétés : signification statistique, valeurs minimale et maximale, compatibilité avec l'ordre (partiel) des courbes de Lorenz, sensibilité à des transferts entre individus, décomposabilité additive.

Parmi ces mesures d'inégalité, le coefficient de Gini est un des plus utilisés pour évaluer les disparités de revenus. De par sa forme fonctionnelle "fortement" non linéaire, il apparaît a priori difficile d'évaluer la précision d'un tel indice quand il est calculé sur des données obtenues dans une enquête par sondage, car la théorie habituelle ne s'applique pas. **Jean-Claude Deville** (DG, unité méthodes statistiques) s'est attaqué à ce problème, et a montré que l'estimation de la variance du coefficient de Gini se ramène à l'estimation de la variance du total d'une variable artificielle linéarisée assez exotique mais facile à construire.

Jérôme Accardo et **Madior Fall** (DG, division revenus et patrimoine des ménages) ont comparé les indicateurs d'inégalité des revenus tirés de différentes enquêtes-ménages réalisées par l'Insee de 1984 à 1993. Utilisant la méthode proposée par J.-C. Deville, ils ont calculé la précision des coefficients de Gini, et ont conclu à l'absence de significativité des évolutions de cet indicateur d'une enquête à l'autre. Ils ont également étudié la robustesse de l'indicateur, en évaluant l'impact de "contaminations" simulant différents types d'erreur de mesure : ces simulations tendent à prouver l'instabilité de l'indicateur, résultats que vient nuancer cependant, selon les auteurs, le constat de la faible dispersion du coefficient de Gini dans les diverses enquêtes, qui reposent pourtant sur des modes de collecte assez différents.

Les statistiques locales

La réponse à la demande de résultats locaux - de plus en plus forte depuis quelques années - peut s'obtenir par deux méthodes : en réalisant une enquête locale, ou en exploitant une enquête nationale par des techniques adaptées. **Sophie Destandau** (DG, unité méthodes statistiques) a dressé un panorama des méthodes d'estimation sur petits domaines à partir d'une enquête par sondage, en précisant les avantages et les inconvénients des différentes méthodes d'estimation : estimations "directes", estimations "synthétiques", estimations "combinées". A titre d'illustration, certaines de ces techniques ont été appliquées pour obtenir des résultats régionaux sur l'enquête Education de 1992, et les performances de ces estimateurs ont pu être comparées.

La France ne disposant pas de registres de population, les recensements de la population y constituent la base du système d'informations socio-démographiques. Cependant, entre deux recensements, l'actualisation de certaines données est nécessaire, notamment à un niveau géographique fin, d'autant plus que les recensements ont tendance à s'espacer. Un groupe de travail a été mis en place à l'Insee, avec pour objectif de proposer une nouvelle méthode d'estimations locales de population. **Jean-Claude Labat** (DG, direction des statistiques démographiques et sociales) a présenté les principales conclusions de ce groupe : réaliser une synthèse efficace et robuste des informations apportées par différentes sources administratives et mobiliser un nombre suffisant de "bonnes" sources, constituant ainsi un système "multi-sources" souple et fiable, sans être trop complexe.

L'Insee a élaboré une nouvelle nomenclature spatiale, le "Zonage en aires urbaines" : elle remplace les Zones de peuplement industriel ou urbain (ZPIU) qui, définies en 1962, couvraient en 1990 les trois quarts du territoire et abritaient 96 % de la population. **Thomas Le Jeannic** (DG, division statistiques et études régionales) a montré en quoi ce nouveau zonage constitue une nouvelle approche territoriale de la ville et de son espace périurbain, dont il donne une image plus restrictive : défini cette fois autour de 361 pôles urbains, sélection des plus grandes unités urbaines, le périurbain doit en outre dépasser un seuil élevé de migrants alternants (40%) vers ces pôles.

Les enquêtes sur des sujets sensibles

Alain Giami et **Alfred Spira** (Inserm) ont présenté un bilan méthodologique de trois enquêtes françaises sur le comportement sexuel : l'enquête ACSF auprès des adultes métropolitains, l'enquête ACSAG auprès des adultes en Guadeloupe, Martinique et Guyane, et l'enquête ACSJ auprès des jeunes de 15 à 18 ans en métropole. La méthodologie de ces enquêtes est bien différenciée : téléphone pour l'enquête ACSF, face à face au domicile des adultes pour l'enquête ACSAG, face à face dans les établissements scolaires ou d'apprentissage pour l'enquête ACSJ. Les méthodes d'échantillonnage employées, toujours probabilistes, s'avèrent également d'une grande diversité (surreprésentation des situations à risque vis-à-vis du sida dans l'échantillon ACSF, grâce à une vaste enquête-filtre). Les intervenant ont également mis l'accent sur les difficultés soulevées par de telles enquêtes : comment interroger quelqu'un sur sa sexualité, en respectant sa sensibilité... en particulier pour les jeunes ? Quels mots employer ? Quel sens donner aux divergences de réponse des hommes et des femmes sur leurs pratiques conjointes ? Une telle recherche a nécessité le travail en commun d'une vaste équipe pluridisciplinaire formée de psychologues, psychosociologues, linguistes, sociologues, économistes, médecins épidémiologistes, démographes et statisticiens !

L'Ined et l'Association l'Elan Retrouvé ont réalisé pendant les hivers 1994-1995 et 1995-1996 deux enquêtes auprès d'un échantillon représentatif de personnes sans domicile utilisatrices de services d'hébergement et de distribution de nourriture, à Paris. **Maryse Marpsat** et **Jean-Marie Firdion** (Ined), et **Viviane Kovess** (Association l'Elan Retrouvé) ont exposé la méthodologie largement commune de ces enquêtes, inspirée des méthodes de sondage et d'estimation d'enquêtes américaines. Il se pose bien sûr des problèmes d'échantillonnage, puisque la population échappe à la base de sondage classique des logements, mais aussi des questions de fond comme d'éthique, ainsi que les difficultés pratiques que l'on peut imaginer. L'enquête de l'Ined, outre sa fonction d'expérimentation pour une enquête à plus grande échelle, cherchait à explorer les processus conduisant à la situation de sans-domicile. L'enquête de l'Elan Retrouvé avait pour objectif de connaître la prévalence des principaux problèmes de santé mentale dans la population des sans-abri, la présence de maladies somatiques sévères et de handicaps, et l'utilisation de soins.