

PRESENTATION

*Ketty Attal-Toubert
Unité méthodes statistiques - Insee*

Organisées par l'unité Méthodes statistiques de l'Insee, les sixièmes Journées de méthodologie statistique¹ se sont tenues les 17 et 18 mars 1998 à Paris, au Centre de conférence Pierre Mendès France du ministère de l'Economie, des Finances et de l'Industrie. Le succès de ce rendez-vous annuel ne se dément pas : au total, environ 400 personnes ont assisté aux différentes conférences et sessions thématiques.

Les questions relatives à la précision des estimations ont occupé une place importante : présentation du logiciel POULPE de calcul de précision des enquêtes, exposé comparatif, par un confrère de Statistique Canada, de différentes méthodes d'estimation de la variance en présence d'imputation de la non-réponse. Les nouvelles orientations relatives aux recensements de population ont été débattues, et un point de vue américain a été donné par un collègue du Bureau of the Census des Etats-Unis. Il a également été question de la méthodologie des enquêtes par panel, avec en exemple le panel européen de ménages, de la collecte par enquêteurs et des indices de prix. Enfin, des enseignants, parmi lesquels un professeur ivoirien, ont rendu compte d'expériences de formation active aux enquêtes.

La conférence inaugurale de ces sixièmes Journées a porté sur les préoccupations européennes en matière d'harmonisation et de comparabilité.

Tant ses producteurs que ses utilisateurs sont de plus en plus attentifs à la qualité de l'information statistique. Dans la conférence inaugurale, Daniel Defays (Eurostat) a expliqué que dans le contexte actuel de globalisation économique et d'intégration européenne, la comparabilité des données, au-delà de leur précision intrinsèque, était une composante essentielle de la qualité de l'information statistique. Une harmonisation complète des méthodes de mesure ne paraît toutefois pas envisageable, les spécificités des pays exigeant souvent des méthodologies adaptées. En matière de comparabilité, les progrès doivent donc résider dans une meilleure compréhension des différentes sources, dans la construction d'une théorie de l'erreur qui permette d'identifier les écarts liés aux méthodes, et dans l'utilisation de modèles d'analyse des résultats permettant de « purifier » les mesures.

1. Les actes des cinq premières Journées sont parus dans la collection *Insee Méthodes*.

Recensements de la population

Lors de la première conférence spéciale, Yves Thibaudeau (Bureau of the Census) a exposé la stratégie américaine, à savoir économie et efficacité, pour le recensement des Etats-Unis de l'an 2000 : développement de partenariats avec des institutions locales et des services de marketing, simplification des opérations et des questionnaires, adoption d'une technologie intelligente (lecture optique par exemple) et d'une méthodologie statistique permettant de corriger les biais liés au sous-dénombrement et à la non-réponse. Le biais de sous-dénombrement devrait être estimé à l'aide d'une technique de « capture-recapture », la correction du biais de non-réponse s'appuiera sur une enquête auprès d'un échantillon de non-répondants.

Dans un exposé introductif au débat sur l'avenir des recensements, Michel Isnard (Insee-DG, département de la démographie) a présenté un projet de recensement en continu, préparé par une mission d'étude sur l'avenir des statistiques de population. Les difficultés que pose un recensement général (coût, adéquation du personnel de l'Insee, acceptation par la population) ont en effet incité l'Institut à réfléchir à une nouvelle forme de recensement tournant. Différents thèmes ont été abordés, comme la définition de la population légale, l'organisation de la collecte, les liens avec le recensement de 1999, la méthodologie d'échantillonnage et la diffusion.

Le débat a donné lieu à des échanges sur les grandes transformations qu'apporterait le passage à un recensement en continu concernant la nature des statistiques produites, leur fréquence, leur précision, les modalités de diffusion...

Estimation de variance en présence d'imputation

Une deuxième conférence spéciale a été consacrée à l'estimation de variance en présence d'imputation de la non-réponse. En effet, l'imputation des non-réponses peut non seulement introduire un biais dans les estimations, mais aussi fausser le calcul de la variance. Eric Rancourt (Statistique Canada) a présenté puis comparé différentes méthodes : imputation multiple, approche assistée d'un modèle, approche en deux phases, technique du Jackknife, Bootstrap, méthode pour hot-deck, méthode d'imputation de tous les cas, et méthode des échantillons balancés répétés. La méthode assistée d'un modèle est incorporée à un logiciel informatique en cours de développement par Statistique Canada.

Le panel européen de ménages

Jean-Claude Deville (Insee-DG, unité Méthodes statistiques) a ouvert cette session par un exposé des contraintes logiques relatives à la construction des données de

panel, contraintes liées à la nature des unités, à leur identification, à la façon de les échantillonner, aux techniques de pondération, à la correction de la non-réponse. Il a introduit un formalisme unificateur des différentes formes de collecte dans le but de rendre accessibles les analyses habituelles, notamment le calcul de variance.

Le panel européen de ménages² est une enquête réalisée depuis 1994 dans l'ensemble des pays de l'Union européenne, à objet d'observer l'évolution de la situation d'activité et de revenus des individus des ménages. Dominique Ansieau (Insee, direction régionale de Lorraine) s'est appuyé sur l'exemple de cette enquête pour évoquer le problème général de l'attrition des panels, et détailler un certain nombre de dispositions pratiques permettant d'atténuer le phénomène : amélioration de l'identification et du suivi (cas de déménagements) des ménages et des personnes, remise de cadeaux aux ménages répondants, réalisation des enquêtes successives auprès d'un même ménage par le même enquêteur.

Christine Chambaz et Nadine Legendre (Insee-DG, division Revenus et patrimoines des ménages) se sont penchées sur l'étude et la correction de la non-réponse dans le panel européen de ménages : analyse des caractéristiques des non-répondants, et correction de la non-réponse par post-stratification. Pour le calcul des pondérations, ont été éprouvées deux hypothèses distinctes sur le comportement des non-répondants. L'influence du choix du modèle sur les statistiques produites à partir de l'enquête semble limitée.

Collecte et enquêteurs

Toujours dans le contexte des enquêtes auprès des ménages, Catherine Berthier et Bernard Neros (Insee-DG, unité Méthodes statistiques) ont présenté une méthode de mesure de l'effet enquêteur, défini comme la part de variation des réponses due au fait que tous les ménages ne sont pas interviewés par le même enquêteur. La méthode, qui a été testée sur la première vague du panel européen, consiste à couper en deux les zones géographiques échantillonnées et à les répartir sur des paires d'enquêteurs. On calcule alors, pour une statistique donnée, la différence entre les deux échantillons, et on la situe sur une distribution empirique, construite par mélange aléatoire des réponses des ménages. Une conclusion est que l'effet enquêteur est variable selon les questions posées, mais ne joue pas sur la non-réponse totale (toutes causes de non-réponse confondues).

La plupart des enquêtes statistiques auprès des entreprises sont encore réalisées par courrier. Mais ce mode de collecte est mal adapté au cas des grandes entreprises, dont la structure ne cesse de se complexifier. Il faut également compter avec le rejet

2. Cf. l'article de Dominique ANSIEAU, Chantal CASES et Christine CHAMBAZ : « Le panel communautaire de ménages », *Courrier des statistiques* n° 83-84, décembre 1997.

accru de la charge administrative, qui obère les taux de réponse. De fait, le recours aux enquêteurs apparaît de plus en plus nécessaire. Chantal de Barry (Insee-DG, cellule Coordination des activités d'enquêtes) et Marcel Perrot (Insee-DG, division Harmonisation des enquêtes auprès des entreprises) ont présenté une expérimentation visant à initier un réseau d'enquêteurs terrain, sur des opérations de faible volume. Le premier bilan est positif : amélioration du taux de réponse, de la qualité des réponses et des relations avec les entreprises. Une expérimentation plus poussée est prévue pour 1998, avant le déploiement complet du dispositif.

Enfin, Philippe Houssay (Insee-DG, pôle Infrastructures géographiques) a exposé les grandes lignes du projet CICN, cartographie infracommunale numérisée : cartographie automatique des districts du recensement, cartographie thématique à l'îlot et géocodage à l'adresse. Une démonstration de cartographie numérisée pour les immeubles de la Réunion a été effectuée.

Le logiciel de calcul de précision POULPE

Nathalie Caron (Unité Méthodes statistiques) a présenté les aspects méthodologiques du logiciel POULPE, programme optimal et universel pour la livraison de la précision des enquêtes, qui permet de calculer des statistiques simples (totaux) ou complexes (ratios ou fonctions de plusieurs variables) ainsi que la variance de ces estimations pour des plans de sondages très divers : sondage aléatoire simple, sondage à plusieurs degrés, en plusieurs phases, prise en compte de la non-réponse ou du calage sur marges.

Jean-Noël Petit (Insee-DG, unité Méthodes statistiques) a ensuite exposé les aspects informatiques de ce logiciel, programmé en SAS. POULPE modélise un sondage complexe par un arbre, dans lequel chaque branche renferme les caractéristiques d'un sondage élémentaire : type de tirage, unités tirées, variables auxiliaires, etc. A partir de ce modèle et des fichiers de données, le logiciel applique les formules de calcul (probabilités d'inclusion et variance) et produit les statistiques demandées avec leur précision.

Dans le cadre de l'exploitation de l'enquête logement, David Le Blanc (Insee-DG, division Logement) a notamment estimé au moyen du logiciel POULPE la répartition des ménages par grand statut d'occupation du logement (propriété, location HLM, location libre) et les flux quadriennaux afférents. Il a pu montrer, par le calcul de l'intervalle de confiance, que la progression apparente de la proportion de propriétaires entre 1992 et 1996 était en réalité une stabilisation. Il a souligné le rôle important que peut jouer POULPE pour inciter les statisticiens à calculer systématiquement la précision des estimations et à la communiquer aux utilisateurs.

Indices

Les substitutions que les consommateurs effectuent entre les produits ou entre les lieux d'achat peuvent être prises en compte dans l'indice des prix à la consommation grâce à une technique de chaînage d'indices. Afin de réduire le délai entre deux chaînages (actuellement un an en France), François Magnien et Jacques Pognard (Insee-DG, division Prix à la Consommation) ont proposé de recourir à l'utilisation des micro-données produites par les sociétés de marketing. Ces données scannées offrent en effet des informations particulièrement précieuses, tant sur les prix que sur les quantités consommées. Une expérience pratique, et concluante, a été menée à partir de données du Panel AC Nielsen relatives au café.

Les services de télécommunication sont en constante évolution, ce qui rend difficile le calcul des purs mouvements de prix. Dans sa communication, Charles Bérubé (Insee, direction régionale des Pays de la Loire) a discuté l'utilisation d'une valeur unitaire pour calculer un indice de prix à la production de ces services. Il a exposé les avantages et les inconvénients d'un indice à valeur unitaire, et les techniques qui peuvent servir au calcul.

François Lequiller (Insee-DG, département des comptes nationaux) a fait le point sur la polémique relative au biais des indices de prix à la consommation. Une commission, dite Boskin, avait en effet annoncé en 1996 que les prix étaient surestimés de 1,1 % par an dans l'indice des prix à la consommation américain. L'Insee a montré qu'en France le biais était de bien moindre ampleur. Mais de grandes lacunes subsistent dans le calcul des indices. L'exploitation des données scannées améliorera la prise en compte des phénomènes de substitution, mais il faudrait également utiliser les régressions hédoniques pour estimer les changements de qualité, et regrouper les services médicaux.

Faire pour apprendre : trois expériences de formation active aux enquêtes

L'Ensaï, Ecole nationale de la statistique et de l'analyse de l'information, a la volonté de compléter les acquis théoriques de ses élèves par la réalisation de projets. Pour les élèves de la filière statistique, il s'agit de réaliser une enquête grande nature, répondant à un problème posé par un commanditaire extérieur. En 1996-1997 par exemple, le commanditaire était l'unité d'économie et de sociologie rurale de l'Inra de Rennes, et l'enquête portait sur les différentes attitudes face au problème de la « vache folle ». Yves Tillé et Michel Simioni (Ensaï) ont décrit les problèmes pédagogiques posés par l'organisation d'un travail collectif de grande dimension à des fins d'apprentissage : distinction entre milieu scolaire et milieu professionnel, difficultés organisationnelles liées au risque d'échec, au degré de mobilisation des

élèves, au travail en équipe... Ils estiment qu'en la matière, le rôle des enseignants recouvre quatre grandes fonctions : ressources et conseil, structuration du travail, arbitrage, et évaluation.

Dans le cadre de leur formation à Libourne, les contrôleurs stagiaires de l'Insee doivent réaliser, sur cinq semaines et demie, une enquête complète proposée par un commanditaire extérieur. Bertrand Roucher (CEFIL) a présenté les objectifs de cette opération : élaborer un système de recueil de données, gérer efficacement une enquête, produire des données normalisées après traitement de l'information, restituer et mettre en forme l'essentiel des informations contenues dans un ensemble de données, rédiger une première analyse et la présenter oralement, organiser et réaliser un travail en groupe. Malgré les écueils rencontrés tels que le choix du partenaire, la détermination du sujet et l'organisation pratique, la première expérience, sans être une réussite parfaite, a montré que les objectifs initiaux pouvaient être atteints. La deuxième expérience est en bonne voie. Il faut espérer que ce type d'opération pourra être renouvelé chaque année.

Benjamin Zanou (Côte d'Ivoire) a décrit le cours de pratique d'enquêtes à l'Ensea, Ecole nationale supérieure de statistique et d'économie appliquée et passé en revue les problèmes que suscite son organisation : recours à du personnel extérieur, encadrement, évaluation des étudiants et perception des enquêtes Ensea par la population. Le point de vue des étudiants est plutôt positif, malgré quelques appréhensions pour aller sur le terrain et des séances de classe parfois houleuses.