

# Synthèse des interventions (\*)

## Conférences spéciales

La première conférence spéciale, par **P. Lavallée** (Eurostat), a traité des méthodes d'échantillonnage pour sélectionner des panels d'entreprises (Eurostat envisageant la création d'un réseau de panels d'entreprises au niveau européen). Le plan de sondage couramment employé pour les enquêtes-entreprises fait usage de l'échantillonnage aléatoire simple (EAS) (après une stratification selon la taille). Dans le cas de panels, certaines complications apparaissent avec l'EAS pour le traitement des "naissances", des "morts", et des changements de strates. Il semble alors préférable d'utiliser l'échantillonnage de Bernoulli, aussi connu sous le nom d'échantillonnage de Poisson, qui simplifie ces traitements et rend possible une certaine rotation de l'échantillon.

La deuxième conférence spéciale, par **M. Hidiroglou** (Statistique Canada), a traité des problèmes d'estimation dans les petits domaines, lorsque les données disponibles ont été recueillies à partir d'un plan de sondage non conçu pour de telles estimations. L'estimateur direct (somme des valeurs dilatées) est alors conditionnellement biaisé et sa variance peut être très élevée. Afin de remédier à ce problème, Statistique Canada utilise toute une gamme de méthodes, qui se servent de modèles liant ces petits domaines à des données auxiliaires provenant de bases de données administratives ou de recensements : estimation "synthétique", post-stratification, méthodes d'estimation dépendant de la taille réalisée de l'échantillon dans le petit domaine, méthodes de régression. L'évaluation de ces méthodes en termes de biais, variance et erreur quadratique moyenne, a été réalisée à partir de deux grandes simulations utilisant des données basées sur une enquête sur les entreprises et une enquête sur les ménages.

## Session 1 : échantillonnages complexes

**F. Cotton** (Direction Générale de l'INSEE, division projets pour les statistiques d'entreprises et les répertoires) a présenté une méthode générale de coordination d'échantillons provenant d'une même base de sondage, fondée sur une technique d'attribution à chaque unité d'un numéro aléatoire et de gestion de ces numéros. Cette méthode permet, dans le cadre des Enquêtes Annuelles d'Entreprises, le contrôle du recouvrement des échantillons de deux enquêtes stratifiées différemment, ou la coordination d'une enquête d'établissements et d'une enquête d'entreprises.

Dans le cadre du tirage d'un échantillon à probabilités inégales, **P. Ardilly** (DG, division des méthodes statistiques et des sondages) a proposé des algorithmes permettant d'assurer une "représentativité approchée" optimale de l'échantillon sur un ensemble de variables auxiliaires connues sur la population entière. La prise en compte de l'information auxiliaire se fait donc au niveau du tirage, et non au niveau de l'estimation (cf session 4), permettant ainsi l'utilisation de l'estimateur simple des valeurs dilatées. Ces techniques, applicables pour des petites tailles d'échantillons dans des populations restreintes, ont été utilisées pour le tirage des unités primaires de l'échantillon-maitre de 1990.

Dans le cadre de la rénovation de l'indice des prix à la consommation, des réflexions ont été menées sur les échantillonnages utilisés pour le calcul de l'indice. **F. Guglielmetti** (DG, division prix de détail) a montré comment on peut optimiser l'échantillon d'agglomérations utilisé pour le calcul de l'indice : on conserve la même précision sur l'indice global en diminuant le nombre d'agglomérations de 105 à 80. D'autre part, on peut agir sur l'échantillon de relevés par catégorie de commune-variété de façon à obtenir, sous des contraintes de coût, des indices de postes "publiables", i.e. de variance inférieure à un certain seuil (tout au moins pour les gros postes).

(\*) déjà paru dans le Courrier des statistiques, n° 57

## Session 2 : précision des données issues d'enquêtes complexes

La procédure de tirage des enquêtes de l'INSEE auprès des ménages dans l'échantillon-maître issu du recensement de la population permet dès le tirage de l'échantillon une estimation anticipée de la précision de statistiques simples (totaux) ou plus complexes sur des variables issues du RP. A partir d'un test utilisant des échantillons de l'enquête Loyers et Charges, **C. Brutel** (DG, division synthèses relatives aux biens et services) a montré la fiabilité de cette méthode, qui sera intégrée dans la chaîne de tirage des échantillons provenant de l'échantillon-maître de 1990.

**N. Roth** (DG, division emploi) a décrit les grandes lignes de l'échantillon de l'enquête emploi qui doit être mis en place en 1992. Des simulations réalisées sur l'échantillon actuel ont permis de mettre en évidence l'influence, sur la précision de quelques grandeurs mesurées à l'enquête, de la taille des aires, de la stratification, de la "régionalisation", du plan de rotation.

## Session 3 : méthodologie des indices

La difficulté de calculer un indice des prix dans les branches à évolution technologique rapide vient de l'impossibilité de définir un panier de produits stable dans le temps. On peut dans ce cas utiliser des "méthodes hédoniques" (ou méthodes de régression) qui supposent que le prix d'un bien est fonction de l'ensemble de ses caractéristiques, selon une relation (linéaire ou autre) que l'on peut estimer par de simples moindres carrés ordinaires : l'indice est alors construit à partir du panier constitué par les caractéristiques. **A. Moreau** (DG, division prix de vente industriels) a présenté une utilisation de ces méthodes dans le cas des microordinateurs (dont les caractéristiques sont le type de processeur, la vitesse d'horloge, la capacité sur le disque dur ...), qui intervient dans le cadre de la mise en place d'un indice de prix de l'informatique.

**J.-P. Berthier** (DG, division comptes et études de l'industrie) a exposé les principales caractéristiques de l'indice du coût de la construction : méthode, échantillonnage, redressements, calcul de l'indice.

Dans l'indice des prix à la consommation, les produits frais (poissons, crustacés, légumes, fruits, fleurs et plantes) font l'objet d'un traitement particulier, en raison de l'offre très variable de ces produits au cours du temps. Dans le cadre de la rénovation de l'indice des prix, une nouvelle méthode de calcul de l'indice des produits frais, présentée par **M. Prime** (DG, division prix de détail), a été élaborée pour remplacer l'ancienne, qui subissait certaines critiques, liées à sa mauvaise prise en compte des mouvements conjoncturels (due à l'utilisation d'une moyenne mobile) et à sa complexité : on chiffre chaque mois un panier mensuel, variable en fonction du mois, et on compare sa valeur à celle du même panier évalué aux prix moyens de l'année de base.

## Session 4 : amélioration d'estimateurs

**L. Berrebi** (DG, division des enquêtes de conjoncture) a montré comment l'utilisation d'une bonne méthode d'estimation peut rapporter gros : dans le cadre de l'estimation du taux d'accroissement annuel de l'investissement, remplacer l'estimateur par le ratio utilisé habituellement par l'estimateur par régression simple (tous deux utilisant l'"information auxiliaire" que constitue l'investissement de l'année n-1) permet de n'enquêter que 5000 entreprises au lieu de 6000, avec une précision au moins aussi bonne, sinon meilleure.

Il est très fréquent que le responsable d'une enquête par sondage veuille "caler" son échantillon, pour un certain nombre de variables jugées importantes, sur des structures fiables connues par ailleurs ; ce redressement s'effectue en remplaçant les poids de sondage des unités enquêtées par des poids ajustés, si possible "proches" des poids précédents. **J.-C. Deville** (DG, division des méthodes statistiques et des sondages) a montré que ce problème de "calage sur marges" rentre dans un cadre plus général : celui de l'estimation du total d'une variable dans une population, lorsque l'on dispose d'une information auxiliaire sous la forme d'un vecteur X de variables mesurées à l'enquête et dont on connaît le total sur la population.

Cette technique de redressement est opérationnelle depuis 1990 pour les enquêtes auprès des ménages. **O. Sautory** (DG, division des méthodes statistiques et des sondages) en a pré-

senté un exemple, qui montre que l'on peut obtenir des distributions de poids ajustés de formes assez diverses selon la façon dont on mesure les écarts entre les poids initiaux et les poids ajustés, ces différentes distributions conduisant de toute façon à des estimateurs ayant approximativement la même variance sur de gros échantillons.

### Session 5 : estimations localisées

**T. Lacroix** (DG, division emploi) a exposé la méthodologie utilisée pour le calcul de taux de chômage à des niveaux géographiques infra-régionaux, qui utilise des données d'enquête (issues de l'enquête emploi) et des données administratives (provenant de l'ANPE), et présenté la qualité des estimations obtenues : erreurs d'estimation liées à la qualité des données ou aux techniques de traitement statistique, ordre de grandeur de la précision des estimations.

**L. Descours** (DG, division statistiques et études régionales) a présenté une méthode d'estimation de populations locales entre deux recensements, qui utilise deux sources :

- la taxe d'habitation fournit une estimation du nombre de résidences principales (par département, tranche d'unité urbaine, Z.P.I.U.)
- l'enquête-emploi donne une estimation du nombre moyen de personnes par ménage (par tranche d'unité urbaine et Z.P.I.U.).

Le taux de croissance annuel du produit (nombre moyen de personnes par ménage X nombre de résidences principales) est appliqué à la population de l'année précédente, et les chiffres sont calés sur l'estimation France entière réalisée par le département de la démographie.

**F. Jeger** (Direction régionale d'Ile-de-France) a montré comment on peut utiliser de l'information auxiliaire pour améliorer la précision d'une estimation issue d'une enquête lorsque l'on s'intéresse à une zone géographique où l'échantillon est faible. La méthode proposée consiste à construire une moyenne pondérée de la valeur observée et de la valeur estimée par une régression sur les variables auxiliaires, les pondérations étant les inverses des variances de ces deux estimateurs. Cette méthode a été appliquée à l'estimation du revenu par ménage dans les bassins d'emploi d'Alsace, les variables auxiliaires étant le revenu par foyer fiscal et la taille des ménages.

### Session 6 : table-ronde sur les problèmes méthodologiques en région

Cette table-ronde a abordé divers aspects des problèmes auxquels doivent faire face les statisticiens travaillant dans les directions régionales de l'INSEE : spécificités régionales, utilisation d'enquêtes nationales au niveau régional, diversité des demandeurs extérieurs à l'INSEE, isolement (dans les petites DR en particulier) ... ; elle réunissait **M. Hannoun** (DR de Bretagne), qui a introduit la discussion par un exposé, **J.-P. Grouthier** (DR des Pays de la Loire), **O. Lefebvre** (DR de Lorraine), **B. Aubry** (DR d'Alsace), **V. Amand** (DG, division des statistiques et des études régionales).

### Répartition des participants aux journées selon leur provenance

<b>INSEE Direction Générale</b>	90
dont:	
Inspection Générale	2
Secrétariat Général	8
Coordination Statistique et Relations Internationales	3
Statistiques Economiques	28
Statistiques Démographiques et Sociales	25
Etudes et Synthèses Economiques	5
Diffusion et Action Régionale	13
ENSAE	5
 <b>INSEE Directions Régionales et CNI</b>	 47
dont:	
Basse-Normandie	6
Bretagne	4
Centre	4
Champagne-Ardenne	4
Ile-de-France	4
 <b>Autres</b>	 35
dont:	
Statistique Canada et EUROSTAT	2
Services Extérieurs des ministères	21
Universités et organismes de recherche	8
Entreprises privées	4