

ANNE-MARIE DUSSAIX

**La rénovation du recensement de la population. Le choix
du mode de sondage dans les grandes communes**

Journal de la société française de statistique, tome 140, n° 4 (1999),
p. 37-39

http://www.numdam.org/item?id=JSFS_1999__140_4_37_0

© Société française de statistique, 1999, tous droits réservés.

L'accès aux archives de la revue « Journal de la société française de statistique » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

LA RÉNOVATION DU RECENSEMENT DE LA POPULATION

Le choix du mode de sondage dans les grandes communes

Anne-Marie DUSSAIX *

Dans les présentations, exposés ou documents concernant la rénovation du recensement, il est constamment fait mention du fait que l'idée essentielle du projet est d'obtenir une évaluation «en continu» de la population, sans augmentation de coût par rapport aux recensements antérieurs.

Le document très clair de Jean Dumais, Philippe Bertrand et Bertrand Kauffmann nous présente dans une première partie les stratégies d'échantillonnage dans les communes de moins de 10 000 habitants et dans les communes de 10 000 habitants ou plus.

Il est difficile de séparer la présentation et la discussion sur le choix du mode de sondage de celles sur la précision des données. Cependant, ma discussion portera davantage sur la méthode d'échantillonnage retenue et les différentes simulations de plans de sondage qui ont été effectuées (celle de Benoît Riandey portant davantage sur la précision des résultats en relation avec les différents niveaux d'utilisation possibles du recensement).

En ce qui concerne *les communes de moins de 10 000 habitants* et d'après le document qui nous est présenté, des simulations utilisant la méthode de tirage d'échantillons équilibrés ont été effectuées pour la région Rhône-Alpes sur des données du Recensement général de la population de 1990. Ces simulations réalisent cinq groupes de communes bien équilibrés sur les variables Nombre total d'hommes, Nombre de résidences principales, Nombre de ménages avec 2 voitures. Le document n'indique rien cependant sur les résultats obtenus pour d'autres types de variables comme la profession. Il est indiqué aussi clairement que l'équilibre n'est pas préservé pour des niveaux géographiques inférieurs (département, aire urbaine, etc). Les implications sur la précision des résultats annuels devraient être précisées.

Par ailleurs, les tests de stabilité indiquent une bonne stabilité de l'équilibre dans le temps au niveau régional.

En ce qui concerne *les communes de 10 000 habitants ou plus*, le plan de sondage qu'il est prévu de réaliser dans une commune donnée est un plan de

* ESSEC, avenue Bernard Hirsch, BP 105, 95021 Cergy-Pontoise Cédex; dussaix@essec.fr

sondage stratifié selon la stratification IRIS 2000 croisé par taille d'immeuble (avec trois tailles d'immeubles). Au sein de chaque strate, sont créés cinq groupes de rotation. Les groupes de rotation d'immeubles seront visités à tour de rôle sur une période de cinq ans. Chaque année, on dressera la liste des logements du groupe de rotation concerné sur lequel on tirera un échantillon au taux de 40 %. Il s'agit donc, dans chaque groupe de rotation d'immeubles, d'un sondage aléatoire simple qui dispersera l'échantillon de logements sur l'ensemble de l'IRIS 2000.

Il est indiqué qu'en « grande commune », on aura la possibilité d'augmenter l'échantillon de ménages jusqu'à couvrir 100 % des logements du groupe de rotation. Les conditions dans lesquelles ceci sera possible ne sont pas précisées.

Des simulations ont été réalisées et concernent une commune de 21 000 habitants, la commune de Sedan. Elles permettent de comparer différents plans de sondage. La majorité des plans testés en comparaison avec le plan proposé pour le RPP sont des plans à deux degrés. Le plan *C* est un plan de sondage à deux degrés où l'unité primaire est l'immeuble. Les plans *D* et *E* sont aussi des sondages à deux degrés dans lesquels les unités primaires sont des grappes de logements contigus. Enfin, le plan *F* est un plan de sondage à deux degrés où les unités primaires sont constituées d'îlots.

Les comparaisons et commentaires sur ces différents plans de sondage prennent bien en compte deux sources d'erreur dans les résultats d'une enquête par sondage :

- les imperfections de la base de sondage ou *erreurs de couverture*. Il est fait en effet l'hypothèse que le répertoire d'immeubles localisés sera mis à jour en continu à partir de permis de construire, de permis de démolir, de fichiers d'abonnés, etc. Un dispositif de repérage de logements a également été mis en place dans les immeubles d'activité, les logements collectifs, et les immeubles neufs (*cf.* paragraphe 2.2). Cependant, le sondage sur liste de logements dans un groupe de rotation une année donnée suppose de dresser la liste des logements de tous les immeubles du groupe de rotation. Comment et quand sera effectué ce prérecensement, c'est-à-dire l'opération de constitution et de mise à jour de la base de sondage des logements dans le groupe de rotation ?
- *les erreurs d'échantillonnage* à travers les simulations effectuées sur les différents plans de sondage. Sur ce dernier point, la méthode proposée des groupes de rotation présente un avantage tout à fait certain.

Cependant, comme le soulignait déjà le Rapport du Comité Scientifique sur la Rénovation du Recensement de la Population (février 2000), la source d'incertitude sur les résultats que constituent les non-réponses doit être davantage prise en compte. Dans l'article de Dumais *et coll.*, au paragraphe 2.4, cet effet est mentionné comme un inconvénient du sondage en grappes : « effet d'entraînement pervers, par lequel le refus d'une unité de la grappe compromet la participation des autres unités de la grappe ». Le Comité Scientifique posait au contraire la question de savoir si un sondage sur liste dans un groupe de logements tel que celui proposé dans le RPP permettait d'assurer l'adhésion de la population et redoutait que le taux de réponse à ce type d'opération soit

plus proche des taux de réponse aux enquêtes obligatoires non aréolaires de l'INSEE.

Dans les simulations effectuées, ne figure aucun plan de sondage en grappes. L'article précise cependant que «des tests sont actuellement en cours pour évaluer les pertes de précision et les coûts de constitution des grappes, si une telle option était retenue». S'agit-il de simuler de réels plans de sondage en grappes avec des niveaux variables de grappes à partir de données de recensement ?

En ce qui concerne la perte de précision dans les sondages à degrés ou en grappes, les résultats du tableau 2 et du tableau 3 donnent les effets de plan de sondage du sondage à 2 degrés par rapport au sondage aléatoire simple. Le plan de sondage E (tableau 3) par exemple affiche un effet de plan de sondage de :

$$(6, 2/4, 5)^2 = 1,90$$

pour la variable «femmes de 20 à 39 ans».

Cela signifie que la taille d'échantillon devrait être multipliée par 1,9 pour obtenir la même précision dans le plan E qu'avec un sondage aléatoire simple. A taille d'échantillon égale, le plan E conduirait à un intervalle de confiance d'amplitude multipliée par 6, 2/4, 5 soit 1,4. Rappelons que, dans l'enquête Emploi de mars 1986¹, la racine carrée de l'effet du plan de sondage en grappes (de taille comprise entre 32 et 48 logements) par rapport au plan de sondage aléatoire simple de même taille pour la variable Nombre d'actifs hommes est de 1,37 (avec un estimateur par les sommes dilatées, sans prendre en compte d'informations auxiliaires).

En *conclusion*, l'article de Dumais *et coll.*, met clairement en évidence l'intérêt du plan de sondage proposé pour le RRP en ce qui concerne la variance d'échantillonnage. Reste cependant, en particulier, la question de savoir quel sera le taux de réponse associé à une telle méthode. Il semble difficile de diminuer à coût constant l'erreur d'échantillonnage sans dégrader le taux de réponse. Est-ce que le taux de réponse ne doit pas être privilégié dans une opération comme celle du recensement rénové? Peut-on atteindre un taux de réponse acceptable à coût constant ?

1. Deville J.-C. et Roth N. (1986). *La précision des enquêtes sur l'emploi*, Économie et statistique, n° 193-194, pp.127-134.