

CARL-ERIK SÄRNDAL

**La rénovation du recensement de la population
: discussions et commentaires**

Journal de la société française de statistique, tome 140, n° 4 (1999),
p. 63-66

<http://www.numdam.org/item?id=JSFS_1999__140_4_63_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

(Discussions et commentaires)

Carl-Erik SÄRNDAL *

Mon point de départ pour ces commentaires est la constatation de Guy Desplanques que la France se trouvera désormais, en ce qui a trait à la méthodologie du recensement de la population (RP), en voie intermédiaire entre l'Amérique du Nord et la Scandinavie. Le RP est toujours effectué en Amérique du Nord (Canada, États-Unis) par énumération traditionnelle, contrastant avec la méthodologie utilisée en Scandinavie (Danemark, Finlande, Norvège, Suède) où l'existence des registres administratifs permet une méthodologie radicalement différente. L'option scandinave serait difficile ou impossible à implanter en France, faute des registres nécessaires.

1. L'évolution de la méthodologie des recensements

Les techniques relatives au RP évoluent dans les autres pays aussi, y compris l'Amérique du Nord et la Scandinavie. Sur une longue période, il s'agit de changements graduels, pourtant importants. Par contre, le Renouveau du recensement de la population (RRP) en France constitue un changement radical par rapport au passé, un changement perçu par certains interlocuteurs comme « une aventure dans l'inconnu ».

2. La solution française, serait-elle « complexe » ?

Le mot « complexe » figure dans la discussion autour du RRP. A mon avis, le RRP est surtout différent, comparativement aux autres pays mentionnés. Par son utilisation de l'échantillonnage, donc l'observation de « moins que la totalité », et par son aspect rotatif, il contraste avec la vue traditionnelle. Selon cette optique, un recensement n'est plus un examen de la totalité. Les deux idées fondamentales, échantillonnage et rotation, expriment le fait qu'à

* Consultant, 2115 Erinbrook Crescent, n° 44, Ottawa, Ontario K1B 4J5, Canada
e-mail : carl.sarndal@home.com

chaque instant l'observation est partielle plutôt qu'en totalité; en revanche on aura des résultats continuellement actualisés sur la population. La solution est astucieuse plutôt que complexe. Les « constructeurs » du plan du RRP ont fait preuve d'une imagination impressionnante.

3. Qui parle échantillonnage parle calcul de l'erreur probable.

Parce qu'il s'agit de sélection et d'estimation (au lieu du dénombrement total), les statisticiens sortent instinctivement leur arsenal probabiliste visant à cerner l'erreur due à l'observation restreinte. Dans cette tradition, on trouve dans l'article de Dumais *et al.* un discours détaillé sur un mode de calcul de la précision des estimations sortant du RRP. Selon le schéma classique, ces auteurs étudient les phases successives d'une enquête par sondage : plan d'échantillonnage, collecte des données, questionnaires, estimateurs à utiliser, la précision de ces estimateurs. Les auteurs donnent des exemples de calcul de précision pour quelques zones géographiques de la France. Dans l'avenir, le calcul de la précision serait assez simple pour les *estimations globales*; cependant, pour les *estimations locales*, une difficulté s'ajoute : comme le notent les auteurs, la situation se présente comme si l'on avait n observations « répondants », et non moins que $4n$ observations « non-répondants » dont les valeurs manquantes seraient comblées par une « imputation massive », ajoutant par ce fait un terme de variance aux estimations du RRP. Encore faut-il préciser qu'un calcul futur de la variance des estimations RRP devra tenir compte également du terme additionnel lié à la non-réponse proprement dite, dont on sait présentement peu, sauf qu'elle ne sera pas nulle.

4. Questions de précision dans les autres pays.

Dans aucun pays le recensement de la population n'est parfait en ce qui concerne la précision. Dans certains pays, on trouve le problème du sous-dénombrement du RP, variable selon les régions du pays. En plus, dans tous les pays, il existe, malgré l'intention d'une « observation de la totalité », un taux de non-réponse non négligeable, et variable lui aussi selon les régions du pays; ce phénomène ajoute aussi un degré d'inexactitude, ou de manque de précision, dans les résultats du RP. Le sous-dénombrement a fait l'objet de débats intenses entre statisticiens, et on a fait appel à des méthodes statistiques avancées pour élucider et cerner ce genre d'erreur dans le RP traditionnel. Pour ce qui est de la non-réponse, d'autres méthodes statistiques avancées entrent en considération; par exemple une méthodologie d'imputation. D'importants progrès en ce domaine ont été réalisés, notamment pour le RP du Canada.

5. Nulle part, un RP «exact».

J'exprime dans les commentaires 3. et 4. un fait dont toute personne initiée est bien consciente : nulle part on n'a un RP «exact». Partout, cependant, les utilisateurs semblent, jusqu'à date, avoir eu confiance, ou suffisamment de confiance, en les résultats du RP. Pour le RRP français, la question de précision entrera en considération directement et explicitement, à cause de l'échantillonnage et de l'imputation massive. Même les utilisateurs les moins avertis savent qu'échantillonnage équivaut à observation partielle, ce qui amène une erreur due à l'échantillonnage. La question de la précision est présente pour les autres pays aussi, malgré l'ambition d'une collecte exhaustive, mais elle se pose de façon indirecte, après le recueil des données. La confiance des utilisateurs tient au fait que la collecte a été exécutée (autant que possible) en totalité; c'est une idée convaincante et rassurante, pour des raisons plus ou moins valides.

6. Les mesures de précision pour le RRP.

La méthodologie prévue pour le RRP assure les moyens d'une production systématique d'*estimations ponctuelles* pour les différentes variables du RRP et pour les différentes zones géographiques de la France, grandes, moins grandes, petites. En plus, par la méthode de Dumais *et al.* et ses extensions, on possédera également les moyens pour une production systématique de la *précision relative* (coefficient de variation) associée à chaque estimation ponctuelle. Les exemples de calcul de coefficient de variation (cv) de Dumais *et al.* indiquent une «bonne précision» pour certaines zones. Le calcul pourra donc se faire; plus importante, à mon avis, est la question de savoir à quoi et à qui servira cette information sur la précision. En principe, il s'agirait d'une grande masse (toutes les zones, toutes les variables) d'indications systématiques de précision à l'aide du cv. Ce calcul montrera une précision dépendant de plusieurs facteurs, notamment la taille de la région en question. Il faut distinguer plusieurs catégories de personnes concernées : les statisticiens de l'INSEE, les utilisateurs des données du RRP et, parmi ces derniers, d'une part ceux qui s'intéressent suffisamment à la précision et pour qui une précision «faible» serait inacceptable, et d'autre part ceux qui accepteront les chiffres (les estimations ponctuelles) tels quels, sans porter attention à une éventuelle indication de précision.

7. Précision jugée insuffisante : quoi faire ?

L'important article de Dumais *et al.* montre un chemin possible pour le calcul de la précision. Ayant mis l'accent sur le calcul de la précision, cet article nous amène immédiatement à une autre question : Que va-t-on faire avec

cette information sur la précision ? Quel rôle jouera une éventuelle production systématique de la précision (le cv) pour les résultats du RRP ? Si ces indications sont produites uniquement pour une discussion interne à l'INSEE, comment les statisticiens, les « producteurs » des informations du RRP, vont-ils réagir ? Y aura-t-il des règles pour supprimer certaines informations issues du RRP, jugées de qualité inadéquate ? Supprimera-t-on tout simplement la publication d'un chiffre dont le cv dépasse un certain seuil, par exemple de 30 % ? Ces règles seront-elles basées sur la valeur du cv, ou bien sur d'autres règles systématiques ou administratives ? A part la question de la précision (mesurée par la variance ou le cv) il y a l'importante question du biais, analysée plus en détail par Benoît Riandey.

8. Points de repère.

Une précision « uniformément bonne » serait évidemment une importante composante de la qualité du RRP. Quelles autres garanties aimerait-on avoir ? Une comparaison intéressante aurait été, évidemment, de comparer la précision du RRP avec la précision de l'ancien RP en France, ou avec la précision réalisée par le RP en Amérique du Nord ou en Scandinavie. Nous ne pouvons pas répondre. Ceci indique la complexité énorme d'une opération qui est pourtant simple à formuler : On veut avoir des informations fiables sur la population d'un pays de 60 millions ou plus d'habitants. Il y a une certaine faiblesse de la part de la méthodologie de la production statistique. Nous ne sommes pas présentement en mesure, à ce stade de nos connaissances, de répondre clairement à la question : Quelle est la précision des informations issues du RP, en France, en Amérique du Nord, en Scandinavie ? Au sujet du RRP, les observateurs hors de la France se posent la question : Est-ce que la méthodologie proposée pour le RRP en France, nettement avantageuse sur certaines dimensions, sera une méthodologie à suivre ailleurs dans le monde ? « Wait and see » est tout ce qu'on peut faire.

9. Le vrai test du RRP.

La difficulté des comparaisons est agaçante. On ne peut pas comparer les pays en ce qui concerne l'utilité pour les utilisateurs du RP de leurs pays respectifs. La vraie valeur du RRP en France ne se mesurera pas de prime abord par un examen du cv des estimations, mais plutôt par l'acceptabilité du RRP auprès du grand public, par le contenu, le niveau de détail, l'actualité de l'information, par la transparence de la méthode, par une absence d'erreurs flagrantes. La valeur que représente l'accès aux informations continuellement actualisées se compare difficilement avec la valeur d'une précision plus ou moins élevée. Si le cv pour une certaine information statistique est de 5 % ou de 20 %, c'est une question non sans importance pour les statisticiens, mais quel serait son intérêt pour la vaste majorité des utilisateurs si l'on est satisfait sur d'autres critères ?