

ÉCHANTILLONNAGE DES NON-RÉPONDANTS ET AUTRES MÉTHODOLOGIES POUR LE RECENSEMENT DES ÉTATS-UNIS À L'AN 2000

Yves Thibaudeau

1. Introduction

La conduite du recensement décennal de la population aux États-Unis fait partie des obligations constitutionnelles des branches exécutive et législative du gouvernement. Cette obligation s'insère dans un mécanisme légal qui donne lieu, à tous les dix ans, à une redistribution des districts congressionaux (équivalents des districts électoraux) entre les états de la fédération. Cette redistribution a pour but de donner aux citoyens une participation au pouvoir aussi démocratique que possible. Idéalement, les rapports entre les nombres de districts congressionaux de deux états sont fidèles aux rapports de tailles entre les populations de ces mêmes états. Le recensement fournit donc les jalons de la redistribution du pouvoir législatif. Il est impératif que le recensement soit sans biais évident ou erreur grossière. En plus, la nation a le droit d'exiger un recensement de qualité qui utilise des méthodes qui sont à la fois efficaces et économiques.

Le dernier recensement décennal de la population (1990) a entraîné des dépenses de près de trois milliards de dollars. En dépit de ce coût élevé, les biais ainsi que d'autres erreurs de mesure ont empiré par rapport à ceux du recensement antérieur. L'avenir semble promettre une conjecture encore plus difficile pour le recensement de l'an 2000.

Cette situation a entraîné une révision profonde des méthodes et des pratiques suivies dans la conduite du recensement décennal. Rappelons-nous qu'à l'époque du dernier recensement, les organes gouvernementaux au niveau fédéral étaient aux prises avec des problèmes de finances déficitaires qui prenaient alors des dimensions politiques. Le vérificateur général a donc procédé à une enquête diligente dans le but d'identifier les fautes et de suggérer des lignes de conduite qui auraient pour effet de mieux contrôler le coût des opérations de recensement, tout en assurant des résultats fiables. Par ailleurs, le congrès des États-Unis décida de parrainer trois commissions d'études, sous l'égide de l'Académie Nationale des Sciences, chacune regroupant plusieurs experts nationaux et internationaux pour en arriver à des

recommandations spécifiques pour corriger les biais et contenir les dépenses. Les recommandations d'une de ces commissions sont publiées par Steffey et Bradburn (1994).

Parmi les recommandations, la proposition d'un échantillonnage des non-répondants est constante. Le contexte dominant est celui d'une économie d'argent importante. Les premières estimations du vérificateur général sont de l'ordre de quatre cents millions de dollars. Ce montant a été révisé à la hausse à cause de divers facteurs, entre autres la situation défavorable à laquelle nous devons faire face sur le marché de l'emploi. En effet, nous prévoyons une compétition vive avec les employeurs qui puisent dans le réservoir de main d'oeuvre où nous nous attendons à trouver nos énumérateurs.

2. Aperçu général de notre stratégie de recensement pour l'an 2000

Le développement logistique du recensement de l'an 2000 s'organise sous quatre thèmes majeurs :

- 1. Partenariat :** Identifier et enrôler des partenaires locaux qui nous aiderons à réussir le recensement.
- 2. Simplicité :** Maintenir une simplicité dans nos opérations, individuellement, et dans l'ensemble.
- 3. Technologie Intelligente :** Acquérir des technologies qui ont des bénéfices immédiats : capture optique digitale, logiciels de couplage.
- 4. Méthodologie Statistique :** Utiliser les méthodologies statistiques pour alléger les opérations, réaliser des économies de fonds, et réduire les biais.

Pour le reste de cette partie, nous révisons les trois premiers thèmes et la façon dont nous les abordons dans nos opérations. Le thème de la méthodologie statistique fait l'objet d'une partie entière.

2.1 Partenariat et marketing

Depuis quelques années nous travaillons à convaincre les institutions locales que c'est dans leur intérêt de coopérer avec nous à la préparation et l'administration d'un recensement de meilleure qualité. Ces nouveaux partenaires sont souvent les gouvernements locaux ou tribaux. Pour le moins ces institutions sont touchées par le recensement, dans la mesure où les octrois de fonds du gouvernement fédéral leurs sont accordés en partie sur la base de la taille de leurs populations. Nous nous

attendons à ce que ces partenaires alertent leurs populations sur l'importance d'une réponse postale au recensement. Nous avons lancé le thème populaire 'Soyez Compté' que, nous l'espérons, les institutions locales vont aider à promouvoir.

Nous avons aussi investi nos énergies à former une alliance avec le service des postes. Celui-ci nous fournira des informations stratégiques d'une valeur inestimable, puisqu'il nous pourvoira des listes d'adresses et des indications quant à l'état de vacance des habitations de chaque quartier. Pour chaque adresse, nous saurons si l'habitation est couramment occupée. Les cas de d'adresses irrésolues sont aussi identifiés.

Finalement, nous avons retenu les services d'une société de marketing privée (Young and Rubicam), dont le contrat est de conduire une campagne de promotion pour la réponse postale au recensement. La diffusion de cette campagne est faite par les médias locaux (télévision, radio, journaux, etc.).

2.2 Simplicité et technologie intelligente

Le thème de la simplicité se traduit par la philosophie de notre approche, plutôt que par une ligne d'action définie. Il en va de même pour le concept d'une technologie intelligente.

Un exemple du souci de simplicité se retrouve dans le questionnaire du recensement. De douze en 1990, le nombre de questions est tombé à sept. C'est le strict minimum requis par la loi. Les questions se rapportent au mode de résidence (propriétaire ou locataire), au sexe, à la race, à l'origine hispanique, et à l'âge. Le questionnaire de papier lui-même a fait l'objet d'un nouveau design, visant à faciliter la navigation d'une question à l'autre, réduisant ainsi les chances d'omission accidentelle de certaines questions.

Par technologie intelligente nous entendons une technologie avancée, mais qui en même temps se doit être strictement fonctionnelle. Par exemple, la capture optique digitale de l'écriture sur les questionnaires de papier est une technologie intelligente. Cette technologie nous permet d'abandonner une partie de l'administration onéreuse des questionnaires de papier. Nous estimons que nous encourageons des bénéfices immédiats substantiels en passant à cette technologie nouvelle.

3. Méthodologie statistique

Le Census Bureau a une tradition de conduire des opérations de contrôle de la qualité sur les recensements décennaux. Ces opérations sont déployées selon des plans d'échantillonnage et des procédures statistiques élaborées. Néanmoins, jusqu'à maintenant les analyses de la qualité ne sont que des fins en elles-mêmes. C'est-à-dire que nous faisons toujours part au public du résultat de nos évaluations, mais nous ne nous servons pas de ces mêmes évaluations pour nous guider dans une correction des biais ainsi mis à découvert.

L'argument classique pour éviter une intervention corrective est que les méthodologies statistiques ne sont pas assez fiables pour garantir qu'une correction d'un biais, comme celui du sous-comptage par exemple, n'introduise pas un nouveau biais, encore plus débilisant que le premier. Jusqu'aux années 70 il est certain que cela tenait du vrai. Par contre, aujourd'hui les méthodologistes, presque unanimement, supportent une intervention technique pour amoindrir certains biais. Le biais du sous-comptage et celui de la non-réponse émergent comme des cibles évidentes. Ce dernier est par ailleurs lié aux questions financières, ce qui ajoute à sa pertinence. Nous avons décidé de mettre sur pied deux opérations d'envergure dans le but de corriger ces deux biais et d'alléger les dépenses.

Pour le reste de cette partie nous allons d'abord revoir de façon sommaire la méthodologie de correction du biais du sous-comptage que nous allons utiliser à l'an 2000. Il s'agit d'une méthodologie de capture-recapture à partir d'un plan d'échantillonnage qui chevauche le recensement. Ces techniques ont subi l'épreuve du temps et ont fait l'objet de nombreuses discussions. Ensuite, nous consacrons le reste de l'article à élaborer notre stratégie d'échantillonnage des non-répondants qui permettra une correction du biais de la non-réponse.

3.1 Corriger le biais du sous-comptage

Pour corriger le biais du sous-comptage il s'agit d'étendre la même méthodologie de contrôle de la qualité que nous avons utilisée dans le passé, mais cette fois-ci nous allons intégrer les estimations du sous-comptage dans les statistiques officielles de la population américaine. L'intégration de nos estimations prennent la forme d'ajustements des comptes de la population. Un des problèmes auxquels nous nous sommes heurtés dans le passé est une conséquence directe de la petite taille de l'échantillon que nous avons alors.

En effet, lors du dernier recensement, pour assurer la stabilité de notre inférence, nous avons dû diviser notre échantillon en strates définies par le degré d'urbanisation des régions qui en faisaient partie. Étant donnée la petite taille de l'échantillon, et pour obtenir une précision que nous jugeons satisfaisante, il fut

nécessaire de regrouper ensemble des régions provenant de plusieurs états distincts (dans la fédération américaine). Le statisticien comprend bien la valeur de créer des strates homogènes du point de vue des variables qu'on doit mesurer. Pour nous, une dissociation géographique ne pose pas de problème. Malheureusement les gouvernements locaux regardent d'un mauvais oeil une pratique qui a recours à des prélèvements démographiques dans des états éloignés des leurs pour ajuster les comptes de leurs populations.

Pour l'an 2000 nous construisons un échantillon de recapture de 750,000 unités d'habitation. Ces unités sont recueillies en 'blocs' qui sont en fait les unités d'échantillonnage. C'est à dire qu'un bloc entier est sélectionné dans l'échantillon. Un bloc, dans une région urbaine, correspond à un pâté de maisons. Dans une région rurale, on essaye de suivre la même définition, de façon plus ou moins arbitraire. Ces blocs sont canevasés méticuleusement, immédiatement après le recensement. En utilisant des techniques de couplage d'enregistrement, nous identifions les citoyens vivant dans ces blocs qui sont recapturés dans l'échantillon. Nous identifions aussi ceux qui sont capturés une fois seulement, soit par le recensement ou soit par la recapture. Ces trois comptes servent ensemble à produire une estimation du biais du sous-comptage. C'est la méthode de l'estimateur duel. Mulry et Spencer (1992) présentent une excellente discussion de cette technique.

Nous soulignons qu'au prochain recensement la taille de l'échantillon de recapture est telle que nous pouvons estimer le biais du sous-comptage des grandes régions métropolitaines avec une précision suffisante, et ce sans avoir recours à une stratification exogène.

3.2 Corriger le biais de la non-réponse

Nous abordons ici un terrain qui n'a pas encore été défriché. À notre connaissance, aucun recensement national n'a eu recours à un plan d'échantillonnage pour corriger les erreurs causées par la non-réponse. Nous avons déjà mentionné que nous sommes poussés vers cette alternative par nécessité. Au dernier recensement, les unités d'habitation qui n'ont pas participé au recensement postal se chiffrent à plus de trente millions, soit plus du quart des unités. Les opérations de récupération des non-répondants sont devenues de plus en plus difficiles à administrer. Les énumérateurs tendent à éviter les quartiers à haut taux de non-réponse, ce qui conduit à une erreur difficile à mesurer. Recruter des énumérateurs efficaces est plus difficile que jamais. La situation à laquelle nous faisons face est un marché de l'emploi clairement défavorable. La rémunération que nous offrons aux énumérateurs est compétitive, mais les termes le sont moins : il est clair au départ que le travail d'énumérateur est pour une période de temps limitée. Par contre, l'énumérateur prospectif peut souvent choisir parmi d'autres emplois plus ou moins précaires, dont certains offrent au moins une possibilité de travail à long terme.

Nous avons décidé de rediriger nos efforts à énumérer seulement qu'une portion des non-répondants. Cette portion doit constituer un échantillon aléatoire et nous devons faire en sorte que la qualité de cette énumération réduite excède celle de n'importe quelle énumération des non-répondants qui se voudrait complète. Nous remplaçons donc une erreur opérationnelle, difficile à mesurer, par une erreur d'échantillonnage, que nous savons bien quantifier.

Notre plan d'échantillonnage ne peut être statique, puisque la population des non-répondants n'est connue qu'après le recensement postal. Nous avons donc arrêté une stratégie bien définie. Dans ce qui suit nous définissons notre projet d'échantillonnage. L'unité géographique cible est la 'tract'. Une tract est un voisinage, elle représente en moyenne 1700 unités d'habitation, soit environ 4000 personnes, mais cette taille peut varier considérablement.

La population des non-répondants d'une tract dont le taux de réponse au recensement postal est moins de 85 % est échantillonnée de la façon suivante :

- l'échantillon est aléatoire,
- la taille de l'échantillon est telle que les répondants au recensement postal, ajoutés aux non-répondants échantillonnés, comptent pour au moins 90% des unités d'habitation de la tract.

La population des non-répondants d'une tract, dont le taux de réponse au recensement postal est égal à, ou est plus élevé que 85%, est échantillonnée de la façon suivante :

- l'échantillon est aléatoire,
- l'échantillon couvre le tiers de la population des non-répondants.

Notre but est de définir une stratégie qui s'explique bien dans un langage que nous voulons accessible au citoyen peu familier avec la statistique. Nous pouvons traduire notre plan de façon simple :

Nous garantissons une énumération avec une erreur réduite relativement aux recensements précédents. Nous assurons une énumération directe d'au moins 90% de la population de chaque voisinage.

Cette stratégie n'est pas parfaite du point de vue technique. Par exemple, puisque les tracts sont de tailles différentes et que le taux de la réponse postale varie lui aussi, nous ne pouvons nous attendre à une erreur d'échantillonnage uniforme. Nous décidons d'abandonner des critères scientifiques plus rigoureux, en espérant qu'une bonne compréhension de notre projet amènera une participation accrue.

4. Résultats et conclusion

Farber et Navarro (1997) rapportent les résultats de simulations de scénarios semblables à celui que nous proposons ici. Leurs résultats sont basés sur le recensement de 1990. Les auteurs simulent des stratégies qui, comme la présente, reposent sur le prélèvement d'un échantillon de recapture ainsi que d'un échantillon de non-répondants, quasi-simultanément. Les auteurs estiment que l'erreur d'échantillonnage totale au niveau de l'état varie entre .2% et .5%, avec une absence théorique de biais. Ces chiffres se comparent avantageusement avec les erreurs encourues au recensement de 1990 : un biais de sous-comptage moyen estimé à 1.6%, et un biais de non-réponse indéterminé, mais non négligeable. Nous rappelons que ce sont précisément les comptes du recensement au niveau des états qui sont à la base de la redistribution des districts congressionaux. Le nombre total de districts congressionaux reste constant. La distribution des districts entre les états seule peut changer, suivant la publication des nouveaux comptes de la population. Une marge de quelques points de pourcentage en plus ou en moins peut signifier un gain ou une perte d'un district. Il est donc très naturel que les milieux politiques s'intéressent à notre stratégie qui utilise des méthodes traditionnelles, aussi bien que des techniques nouvelles. Notre vœu le plus sincère est que cette nouvelle attention portée sur notre méthodologie suscite une participation au recensement plus élevée de la part du public.

Bibliographie

FARBER, J. et NAVARO, A.(1997) : « A Comparison of Alternative Sampling Methodologies for Census 2000 », *1997 Proceedings of the Section on Survey Research Methods, American Statistical Association*, à paraître.

MULRY, M.H. et SPENCER, B.D.(1991) : « Total Error in PES Estimates of Population(with discussion) », *Journal of the American Statistical Association*, 86, 839-863

STEFFEY, D.L. et BRADBURN, N.M.(Eds.)(1994) : *Counting People in the Information Age*, Panel to Evaluate Alternative Census Methods, Committee on National Statistics, National Research Council. Washington, D.C. : National Academy Press.