

# Faut-il ajuster les données de recensement? TO ADJUST OR NOT TO ADJUST? ES NECESARIO AJUSTAR LOS DATOS DEL CENSO?

Nathan Keyfitz

Volume 18, numéro 2, automne 1989

Sous-dénombrement et estimation de population

URI : <https://id.erudit.org/iderudit/010015ar>

DOI : <https://doi.org/10.7202/010015ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Association des démographes du Québec

ISSN

0380-1721 (imprimé)

1705-1495 (numérique)

[Découvrir la revue](#)

Citer cet article

Keyfitz, N. (1989). Faut-il ajuster les données de recensement? *Cahiers québécois de démographie*, 18(2), 247–256. <https://doi.org/10.7202/010015ar>

Résumé de l'article

On oublie trop facilement de se demander quelle est la population visée par les programmes de transferts, et quel était l'objectif de la législation qui les a créés. Les fonds versés aux provinces ou aux États au titre de la santé sont destinés aux personnes malades dépourvues des moyens de se faire traiter, non aux gens en bonne santé ou assez riches pour se payer des soins. Quand on estime une population cible à partir de l'effectif de la population, l'écart entre celui-ci et la population cible atteint des ordres de grandeur beaucoup plus importants que l'écart entre la population dénombrée par le recensement et la population réelle, il faut faire porter nos recherches sur la population cible plutôt que sur l'erreur censitaire, qui est négligeable.

## **Faut-il ajuster les données de recensement ?**

Nathan KEYFITZ\*

Le Canada bénéficie de trois siècles de pratique des opérations de recensement. Depuis cent cinquante ans, ses habitants sont énumérés tous les dix ans et, plus récemment, tous les cinq ans. En dépit de cette longue expérience, les recensements canadiens demeurent entachés d'erreurs et d'omissions. La mise au point de méthodes plus sophistiquées, associée à l'accroissement des mesures de précaution et des budgets, tend à réduire la somme des erreurs, mais rien n'indique qu'on parviendra à les éliminer complètement. En fait, la qualité des dénombrements risque même de diminuer à cause de l'importance que prend la mobilité à notre époque. Par contre, les enquêtes modernes et les analyses démographiques révèlent le sens et l'ordre de grandeur des erreurs, de sorte qu'il est possible de les corriger (Freedman et Navidi, 1986; Romaniuc, 1988).

### **L'AJUSTEMENT DES DONNÉES PEUT AMÉLIORER LE RECENSEMENT**

Un entrepreneur forestier peut évaluer d'un coup d'oeil la quantité de bois contenue dans chaque parcelle d'une concession; l'ayant fait, il choisira un petit échantillon de parcelles, dont il comptera et mesurera soigneusement les arbres, afin de valider son approximation première. Ce genre de démarche existe dans plusieurs domaines. Ne peut-on pas, alors, contrôler la qualité du recensement à l'aide d'une

---

\* International Institute for Applied Systems Analysis, Laxenburg (Autriche).

énumération de petite envergure, mais d'une précision plus grande ?

De nombreux statisticiens débattent la question depuis des années sans parvenir à un consensus. Je crois que la balance penche en faveur du non-ajustement. Les arguments invoqués à l'appui de cette option relèvent en partie de la statistique, mais sont également liées aux attentes de la population et à des politiques d'ordre national.

Au plan statistique, l'ajustement se justifie par l'existence de programmes de transferts aux différentes régions du pays (Keyfitz, 1979; Fellegi, 1980). Il peut s'agir de subventions gouvernementales aussi bien que de représentation au parlement. En combinant l'utilisation d'un échantillon représentatif et la prise en compte des faits démographiques, il est incontestablement possible d'obtenir une évaluation de la population totale du pays plus précise que le décompte fourni par le recensement. Mais, en général, ce n'est pas une donnée aussi globale que l'on cherche. En effet, les programmes fédéraux prévoient l'attribution de fonds à l'échelon infranational : municipalités, régions, provinces, etc. Au niveau des petits ensembles, il est certain que l'amélioration du comptage de la population pour les uns ira de pair avec une détérioration pour les autres.

La tension résultant d'un gain de qualité dans la majorité des cas et d'une perte dans un petit nombre d'autres cas n'est pas facile à dénouer. Si 99 % des unités enregistrent un progrès et seulement 1 % un recul, on sera en général satisfait de l'ajustement. Par contre, si ces proportions sont respectivement de 60 % et de 40 %, on jugera la correction inacceptable. Quelle sera en fait, pour une méthode donnée, la distribution des unités eu égard à l'effet de la correction ? Personne ne peut le dire. Et même s'il existait des moyens statistiques de partager les entités territoriales selon que la méthode utilisée a amélioré ou empiré les données du recensement les concernant, la correction serait-elle davantage acceptable ? La plupart des gens croiront que l'utilisation d'une méthode traditionnelle qui produit une certaine erreur est meilleure qu'un calcul griffonné au dos d'une enveloppe et comportant une erreur de même importance. La question qu'il faut poser à propos des méthodes est celle de savoir combien de cas de l'une peuvent compenser un nombre donné de cas de l'autre. Ce type de considération ne relève pas de l'analyse statistique; il fait intervenir la légitimité des

nombres, leur crédibilité et leur acceptabilité par le grand public.

Il faut garder à l'esprit que l'effectif de la population est une approximation grossière de la population que visent les programmes d'aide. Si l'on organise une cantine dans une école, la clientèle cible sera constituée des élèves dont la mère n'est pas en mesure de préparer le repas de midi. Si l'on songe à des pensions pour les personnes âgées, on visera celles que l'absence de ce secours accule à l'indigence. Un programme d'assurance-santé sera, de même, destiné aux malades qui n'ont pas les moyens de se faire soigner. Dans tous ces cas, le chiffre de la population est certes en corrélation avec celui de la population cible, mais souvent faiblement.

Prenons une population cible dont les chiffres du recensement sont un indicateur, et supposons une corrélation de 0,75 — plus forte donc que celle qui est généralement observée — entre l'indicateur et la cible. Supposons également que l'indicateur est sujet à une erreur aléatoire, avec un écart type de 10 %. Il nous faut alors réduire l'erreur pour améliorer l'indicateur, et porter sa corrélation avec la population cible de 0,75 à environ 0,76. Le gain de qualité est réel, certes, mais pas tellement spectaculaire. (Ce point sera repris plus bas, en termes de variance.)

Mais il ne faut pas oublier qu'il a fallu, pour obtenir cette amélioration, sacrifier le principe selon lequel le recensement doit être établi sous forme de liste nominative mentionnant chaque individu membre de la population.

Il est vrai que dans tout recensement on doit prendre des raccourcis, sans quoi on ne terminerait jamais l'opération. Quand l'agent recenseur est revenu, jusqu'à cinq fois, sonner à une porte sans obtenir de réponse, qu'il a, après enquête auprès des voisins, appris qu'un couple habite derrière cette porte, et qu'il a peut-être même obtenu une indication de l'âge des conjoints, on clôt le dossier de ce ménage. Cela signifie que des informations erronées ou douteuses sont collectées pour ce ménage, c'est-à-dire que deux personnes du voisinage sont comptées deux fois pour compenser l'omission. Cette extension du concept d'énumération est inévitable, mais on n'en abuse pas, l'idéal demeurant la déclaration individu par individu.

On cherche aussi à faire en sorte que les données soient comparables d'un recensement à l'autre, même si on ne fait pas chaque fois le même effort pour que le résultat soit complet. Le grand public comprend ce qu'est un recensement, si l'on ne tient

pas compte des raffinements inhérents à la définition de l'objet du comptage (à supposer qu'il y ait là quelque subtilité !) : les touristes ne doivent pas être dénombrés alors que les résidents temporairement absents doivent l'être. La différence entre un touriste et un résident est facile à comprendre. Par contre, les mesures sophistiquées proposées pour l'ajustement des données ne sont pas à la portée de tous.

Mais surtout, quand on procède à la correction du recensement, on perd l'unicité du résultat. Toute personne au fait des méthodes d'ajustement et intéressée, pour une raison ou pour une autre, à apporter une correction quelconque aura tout loisir de faire preuve d'imagination et d'initiative : l'initiative, dans ce domaine, peut représenter des gains de l'ordre de plusieurs millions de dollars. Cependant, les agences statistiques officielles se verront contraintes, non plus seulement de défendre en cour les données du recensement contre des estimations non contrôlées, mais, tâche plus difficile, de défendre un ajustement particulier, en l'occurrence le leur, contre celui des autres. Un juge d'une cour locale pourra fort bien rendre jugement en faveur d'un plaignant issu du milieu local, mais l'agence centrale de la statistique pourra porter la cause devant des tribunaux supérieurs... Avec des centaines de poursuites de ce genre sur les bras, l'agence sera très occupée.

La correction des données du recensement par l'organisme statistique officiel ouvre une boîte dans laquelle beaucoup trop de personnes aux intérêts différents veulent jeter un coup d'oeil.

### **POURQUOI LE DÉNOMBREMENT EST-IL SI DIFFICILE ?**

S'il y a un certain nombre de personnes dans une pièce, il est certainement possible de fermer les issues assez longtemps pour s'assurer que personne n'entre ni ne sorte, et de faire un décompte dont nul ne pourrait contester l'exactitude. Du point de vue de la masse, Statistique Canada devrait pouvoir procéder ainsi à l'échelon national et compter tous les résidents canadiens jusqu'au dernier. Mais l'analogie est fautive. Le pays n'est pas fermé. Concrètement, le comptage ne peut pas se faire en un instant unique. Distinguer ceux qui ne doivent pas être comptés pose des difficultés. Certaines personnes souhaitent échapper au recensement et ont largement la possibilité de le faire. À notre époque, la plupart

des gens savent lire et écrire et peuvent répondre à une sollicitation écrite, mais la multiplication des enquêtes et l'augmentation de la mobilité annulent cet avantage.

À Statistique Canada, on a fait preuve d'ingéniosité pour contourner les écueils. On expédie le formulaire de recensement par la poste et ainsi, à un coût relativement faible, on obtient l'information désirée de la part de tous ceux qui sont à la fois alphabétisés et prêts à collaborer. L'agence peut alors consacrer une plus grande part de ses ressources aux catégories plus difficiles à énumérer : les analphabètes, les sans-logis, les absents temporaires et les personnes qui ne sont pas disposées à collaborer (par exemple les immigrants illégaux). Ce surcroît d'efforts et de savoir-faire a réduit le sous-dénombrement dans les recensements récents (le dernier mis à part).

Mais il y a des limites à respecter. Au-delà d'un certain seuil, l'appel au sens civique des résidents aura pour résultat que certains seront comptés deux fois. Puisque la vérification des doubles comptes est difficile à mettre en place, il est risqué de trop insister auprès des citoyens sur le fait que, pour le milieu auquel ils appartiennent, il importe qu'ils soient comptés. À titre d'illustration, on a estimé que le recensement américain de 1970 comportait un sous-dénombrement de 1 %; comme le sous-dénombrement était particulièrement élevé pour les minorités et conduisait de ce fait à des inégalités, on a consenti en 1980 un effort spécial et très coûteux, y compris un grand battage publicitaire, et il en est résulté une surestimation de 0,25 % de la population de droit.

### **L'ERREUR DE RECENSEMENT PAR RAPPORT À L'ERREUR SUR LA CIBLE**

On oublie facilement que l'erreur de couverture est dépassée par une erreur différente, liée à l'utilisation du recensement à des fins particulières. Si  $t$  est l'objectif visé (le besoin réel d'une région donnée, soit les repas à l'école dans une partie du pays) et si  $c$  (le sous-ensemble de la population nationale qui habite cette région selon le recensement) est utilisé pour estimer  $p$  (la population réelle), alors la population cible est donnée par la population dénombrée ( $c$ ), plus l'écart entre la population totale et la population cible ( $t-p$ ), plus l'écart entre la population réelle et la population dénombrée ( $p-c$ ). Appelons ( $t-p$ ) l'erreur intrinsèque ou erreur sur la cible, et ( $p-c$ ) l'erreur de couverture. Nous avons alors :

$$t = c + (t - p) + (p - c).$$

Bref, la population cible est la population recensée plus l'erreur sur la cible et l'erreur de couverture. Nous pouvons transposer, élever au carré et transposer encore pour obtenir

$$\sigma^2 (t - p) = \sigma^2 (t - c) - \sigma^2 (p - c)$$

dans l'hypothèse de l'indépendance des deux types d'erreur, de manière à obtenir finalement

$$\sigma (t - p) = \sqrt{\sigma^2 (t - c) - \sigma^2 (p - c)}.$$

La taille relative des deux types d'erreur est très importante lorsqu'il s'agit de déterminer si l'ajustement est raisonnable. Si la population correspond exactement à la cible, alors l'erreur de couverture constitue une véritable injustice. Mais on peut difficilement soutenir que la population cible correspond à la population. Dans les secteurs aisés, il n'est pas nécessaire de servir des repas gratuits dans les écoles, mais en milieu pauvre le besoin avoisine 100 %. D'un autre côté, même si le sous-dénombrement de la population est souvent de l'ordre de 2 % à 3 %, il dépasse très rarement 10 % ou 20 %. Posons que l'erreur liée à la cible est de 50 % et que la sous-énumération est de 10 %.

Le coût de la collecte de données supplémentaires et de leur traitement en vue de la correction peut représenter plusieurs millions de dollars. Mais on ne parviendra pas à un dénombrement parfait : on peut seulement réduire l'erreur d'une proportion inconnue. Nul ne saurait dire si l'ajustement ainsi effectué réduit l'erreur de couverture de moitié, mais on peut affirmer qu'il n'a aucun effet sur l'erreur liée à la cible.

La variance de la somme de deux variables aléatoires indépendantes est la somme des carrés de chacune des deux variables. Si l'erreur type de la vraie population considérée comme estimation de la cible est de 49 % de la moyenne et celle du recensement considéré comme estimation de la vraie population de 10 %, et si les erreurs sont indépendantes l'une de l'autre et de la population, alors l'erreur totale, quand le recensement est pris comme mesure de la cible, est la racine carrée de 0,49 au carré plus 0,1 au carré, c'est-à-dire 0,50. Mais le recensement ne peut être parfait. Supposons que son erreur soit réduite de moitié. Alors, par rapport à la cible, en corrigeant le recensement, nous passerons d'un écart de 50 % à un écart de 49,5 %.

Le tableau 1 présente diverses combinaisons d'erreur sur la cible et d'erreur de couverture et donne, pour chaque combinaison, l'erreur sur la cible du recensement corrigé, exprimée en coefficient de variation. Le tableau suppose que l'ajustement est parfait; dans la mesure où il ne l'est pas, l'amélioration du recensement ajusté est réduite. L'ajustement ne constituera une amélioration sensible que si l'erreur de couverture représente une proportion importante de l'erreur sur la cible.

Ainsi, si l'erreur sur la cible était de 0,1 et qu'elle était complètement due à l'erreur censitaire, alors seulement la correction permettrait d'atteindre exactement l'objectif et de calculer l'allocation budgétaire correspondante. Si l'erreur sur la cible est de 0,2 et le sous-dénombrement de 0,8, l'ajustement portera l'erreur à 0,18, mais un tel gain semble plutôt irréaliste.

TABLEAU 1  
*Erreur type du recensement<sup>a</sup> corrigé utilisé  
comme estimation de la population cible*

| Erreur<br>du recen-<br>sement :<br>$\sigma$ (p-c) | Erreur du recensement c utilisé comme<br>estimation de la population cible : $\sigma$ (t-c) |       |       |       |       |
|---|---|-------|-------|-------|-------|
|   | 0,100   | 0,200 | 0,300 | 0,400 | 0,500 |
| 0,000   | 0,100   | 0,200 | 0,300 | 0,400 | 0,500 |
| 0,020   | 0,098   | 0,199 | 0,299 | 0,399 | 0,500 |
| 0,040   | 0,092   | 0,196 | 0,297 | 0,398 | 0,498 |
| 0,060   | 0,080   | 0,191 | 0,294 | 0,395 | 0,496 |
| 0,080   | 0,060   | 0,183 | 0,289 | 0,392 | 0,494 |
| 0,100   | 0,000   | 0,173 | 0,283 | 0,387 | 0,490 |

a  $\sigma$  (t - p). Exprimée sous forme de coefficient de variation.

La solution qui consiste à réduire partiellement la plus petite source d'erreur sans tenir compte de la source principale n'est pas optimale. C'est comme essayer de mesurer un meuble en plaçant minutieusement le ruban à une extrémité et en le lançant mollement vers l'autre extrémité. Plus spécifiquement, c'est prendre la lettre de la loi où la population est mentionnée, sans se référer au texte qui précise l'objectif de cette loi. Procéder ainsi, c'est être légaliste, perdre l'objectif de vue.

### **L'AUSTRALIE AJUSTE, LES ÉTATS-UNIS NE LE FONT PAS**

Le Bureau of the Census des États-Unis prévoit mener une très importante enquête postcensitaire (30 000 ménages) le

1er juin 1990 ou dans les jours qui suivent. Il appariera alors les mentions nominatives de l'enquête et celles du recensement et mesurera la proportion des omissions. La démarche qui devrait suivre serait d'accroître les données du recensement, pour chaque région, en fonction de la fraction que représentent les individus omis. L'échantillon, considérable à l'échelle nationale, sera petit au niveau local et comportera une erreur échantillonnale élevée. Mais l'erreur de jumelage est plus importante que l'erreur d'échantillonnage. Un petit nombre de noms mal orthographiés peut correspondre à une part importante de la correction. Supposons que 95 % des mentions nominatives de l'enquête aient été retracées dans le recensement pour un territoire donné et que, pour encore 4 % des mentions, le jumelage des mentions soit incertain. La personne «A» de l'enquête post-censitaire et la personne «B» du recensement pourraient être soit la même personne (dont le nom aurait été orthographié différemment), soit deux personnes distinctes. Alors, tout ce que l'on sait vraiment, c'est que le sous-dénombrement se situe quelque part entre 1 % et 5 %. Un résultat qui ne représente que 4 % d'incertitude en regard du chiffre du recensement en constitue 400 % en regard du sous-dénombrement.

Ce genre de considérations a poussé le *Bureau of the Census* à décider de ne pas corriger le recensement de 1990. Ses motifs n'étaient ni strictement techniques ni propres aux États-Unis. De l'avis de Robert Orther, Sous-secrétaire au Commerce, «La correction crée plus de problèmes qu'elle n'en résout [...] Elle sème le doute dans les esprits en ce qui concerne la fiabilité et l'intégrité du recensement et du système statistique fédéral» (notre traduction). Au cours de l'été 1989, après un long débat, le *Bureau* a cependant dû accepter que le prochain recensement puisse être sujet à correction.

De son côté, l'Australie a corrigé les données des recensements de 1976, 1981 et 1986 pour tenir compte du sous-dénombrement, en combinant l'enquête post-censitaire et l'analyse démographique. Elle a estimé que 1,8 % de la population avait été omise dans le recensement de 1986 et augmenté les chiffres dans une mesure équivalente.

Les pressions en faveur de la correction aux États-Unis sont liées à la surreprésentation des minorités parmi les personnes omises. Les municipalités qui comptent de fortes proportions de Noirs et d'hispanophones sont pénalisées par la sous-énumération différentielle de ces groupes. La population du

Canada est loin de comprendre une aussi grande part de personnes incapables de lire ou d'écrire concentrées dans certaines villes. La nécessité de corriger se fait sentir d'autant moins. Or, si l'ajustement ne convient pas pour les États-Unis, il convient encore moins pour le Canada.

La relation entre l'objectif ou cible  $t$ , la population réelle  $p$  et la population énumérée  $c$ , qu'on peut exprimer sous la forme

$$t - c = (p - c) + (t - p),$$

semble sous-jacente au débat sans être explicitée. Si l'on ne tient pas compte, en première approximation, des interrelations entre ces variables, la variance totale (si on utilise le recensement comme estimation de la population cible) est la somme des deux autres variances, et dans la plupart des cas l'erreur de couverture sera de loin la plus petite. Pour cette raison, et parce que ce sont les carrés des erreurs qui sont additifs, la réduction du terme  $(p-c)$  aura un effet insignifiant sur l'erreur totale  $(t-c)$ .

### RÉFÉRENCES BIBLIOGRAPHIQUES

- FELLEGI, I.P., 1980. «Should the Census Count Be Adjusted for Allocation Purposes—Equity Considerations». In *Proceedings of the 1980 Conference on Census Undercount*, U.S. Bureau of the Census, 193-203.
- FREEDMAN, D.A. et W.C. NAVIDI, 1986. «Regression Models for Adjusting the 1980 Census». *Statistical Science*, 1, 3-39.
- KEYFITZ, N., 1979. «Information and Allocations: Two Uses of the 1980 Census». *The American Statistician*, 33, 45-50.
- ROMANIUC, A., 1988. «A Demographic Approach to the Evaluation of the 1986 Census and the Estimates of Canada's Population». *Survey Methodology*, Statistics Canada, 14, 2, 157-171.

**RÉSUMÉ — SUMMARY — RESUMEN****KEYFITZ Nathan — FAUT-IL AJUSTER LES DONNÉES DU RECENSEMENT ?**

*On oublie trop facilement de se demander quelle est la population visée par les programmes de transferts, et quel était l'objectif de la législation qui les a créés. Les fonds versés aux provinces ou aux États au titre de la santé sont destinés aux personnes malades dépourvues des moyens de se faire traiter, non aux gens en bonne santé ou assez riches pour se payer des soins. Quand on estime une population cible à partir de l'effectif de la population, l'écart entre celui-ci et la population cible atteint des ordres de grandeur beaucoup plus importants que l'écart entre la population dénombrée par le recensement et la population réelle. Il faut faire porter nos recherches sur la population cible plutôt que sur l'erreur censitaire, qui est négligeable.*

**KEYFITZ Nathan — TO ADJUST OR NOT TO ADJUST?**

*Insufficient attention has been given to the target of the transfer scheme, the objective sought by the original legislation. Money allocated to states or provinces for health programs is intended to aid those who are sick and lack the means to secure treatment; people who are in perfect health, and those who are rich enough to command treatment, are not the target of the legislation. When population is used as a proxy for such a target the departure of proxy from target is orders of magnitude larger than the departure of the census population from the true population. Research is needed on the target rather than on the trifling error of the Census.*

**KEYFITZ Nathan — ¿ES NECESARIO AJUSTAR LOS DATOS DEL CENSO?**

*Nos olvidamos muy fácilmente preguntarnos cual es la población apuntada por los programas de transferencias y cual es el objetivo de la legislación que los originó. Los fondos entregados a las provincias o a los Estados para la salud son destinados a las personas enfermas privadas de los medios para hacerse tratar, y no a la gente gozando de buena salud o bastante rica para pagarse tratamientos. Cuando evaluamos una población visada a partir de la población efectiva, la diferencia entre ésta y la población visada toma proporciones mucho más importantes que la diferencia entre la población contada en el censo y la población real. Hay que apoyar entonces nuestras investigaciones sobre la población visada en lugar de considerar el error censitario, el cual es despreciable.*