Nº 12-001-X au catalogue
ISSN 1712-5685

Techniques d'enquête

Théorie et méthodologie des enquêtes par sondage : orientations passées, présentes et futures

par J.N.K. Rao et Wayne A. Fuller

Date de diffusion : le 21 décembre 2017

Statistique Canada Statistics Canada

Canada

Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.gc.ca.

Vous pouvez également communiquer avec nous par :

Courriel à STATCAN.infostats-infostats.STATCAN@canada.ca

Téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros sans frais suivants :

- Service de renseignements statistiques 1-800-263-1136
- Service national d'appareils de télécommunications pour les malentendants 1-800-363-7629
- Télécopieur 1-877-287-4369

Programme des services de dépôt

- Service de renseignements 1-800-635-7943
- Télécopieur 1-800-565-7757

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site www.statcan.gc.ca sous « Contactez-nous » > « Normes de service à la clientèle ».

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, les entreprises, les administrations et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

Signes conventionnels dans les tableaux

Les signes conventionnels suivants sont employés dans les publications de Statistique Canada :

.	indisponible pour toute période de référence
..	indisponible pour une période de référence précise
...	n'ayant pas lieu de figurer
0	zéro absolu ou valeur arrondie à zéro
0^s	valeur arrondie à 0 (zéro) là où il y a une distinction importante entre le zéro absolu et la valeur arrondie
^{p}	provisoire
^{r}	révisé
x	confidentiel en vertu des dispositions de la *Loi sur la statistique*
^{E}	à utiliser avec prudence
F	trop peu fiable pour être publié
*	valeur significativement différente de l'estimation pour la catégorie de référence (p<0,05)

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2017

Tous droits réservés. L'utilisation de la présente publication est assujettie aux modalités de l'entente de licence ouverte de Statistique Canada.

Une version HTML est aussi disponible.

This publication is also available in English.

Techniques d'enquête, décembre 2017
Vol. 43, nº 2, pp. 159-176
Statistique Canada, nº 12-001-X au catalogue

Théorie et méthodologie des enquêtes par sondage : orientations passées, présentes et futures

J.N.K. Rao et Wayne A. Fuller[1]

Résumé

L'exposé retrace l'évolution de la théorie et de la méthodologie des enquêtes par sondage au cours des 100 dernières années. Dans un article fondamental publié en 1934, Neyman jetait les bases théoriques de l'approche axée sur l'échantillonnage probabiliste pour l'inférence à partir d'échantillons d'enquête. Les traités d'échantillonnage classiques publiés par Cochran, Deming, Hansen, Hurwitz et Madow, Sukhatme, ainsi que Yates au début des années 1950 étendaient et étoffaient la théorie de l'échantillonnage probabiliste, en mettant l'accent sur l'absence de biais, les caractéristiques exemptes de modèle, ainsi que les plans de sondage qui minimisent la variance selon un coût fixe. De 1960 à 1970, l'attention s'est portée sur les fondements théoriques de l'inférence à partir de données d'enquêtes, contexte dans lequel l'approche dépendante d'un modèle a suscité d'importantes discussions. L'apparition de logiciels statistiques d'usage général a entraîné l'utilisation de ces derniers avec des données d'enquêtes, d'où la conception de méthodes spécialement applicables aux données d'enquêtes complexes. Parallèlement, des méthodes de pondération telles que l'estimation par la régression et le calage devenaient réalisables et la convergence par rapport au plan de sondage a remplacé la contrainte d'absence de biais comme critère pour les estimateurs classiques. Un peu plus tard, les méthodes de rééchantillonnage gourmandes en ressources informatiques sont également devenues applicables à des échantillons d'enquêtes à grande échelle. L'augmentation de la puissance informatique a permis des imputations plus avancées des données manquantes, l'utilisation d'une plus grande quantité de données auxiliaires, le traitement des erreurs de mesure dans l'estimation, et l'application de procédures d'estimation plus complexes. Une utilisation marquante de modèles a eu lieu dans le domaine en expansion de l'estimation sur petits domaines. Les orientations futures de la recherche et des méthodes seront influencées par les budgets, les taux de réponse, le degré d'actualité des données, les outils améliorés de collecte des données et l'existence de données auxiliaires, dont une partie proviendra des « mégadonnées ». L'évolution des comportements culturels et de l'environnement physico-technique aura une incidence sur la façon de réaliser les enquêtes.

Mots-clés : Collecte des données; histoire de l'échantillonnage; échantillonnage probabiliste; inférence à partir d'enquêtes.

1 Introduction

Le présent article a été préparé à la demande de M. Danny Pfeffermann, Ph.D., président de l'Association internationale des statisticiens d'enquête en 2015, qui a proposé le titre ambitieux. L'article a été présenté à l'assemblée de l'Institut international de statistique à Rio de Janeiro, au Brésil, en 2015.

Le titre définit un domaine beaucoup trop vaste pour pouvoir le traiter complètement en un seul article. De surcroît, plusieurs articles de synthèse abordent les sujets qu'évoque ce titre, dont Kish (1995), Bellhouse (2000), Rao (2005), Bethlehem (2009), Brick (2011), Groves (2011) et Brewer (2013). Nous nous inspirons de ces articles dans notre discussion, mais sans chercher l'exhaustivité. Nous donnons une brève évaluation des trois sujets et projetons un certain nombre de situations courantes dans l'avenir. Notre but est de susciter de nouvelles discussions, surtout au sujet des orientations futures. Au-delà de la discussion des controverses concernant l'échantillonnage par choix raisonné, nous nous concentrerons sur l'échantillonnage probabiliste. Le domaine de l'échantillonnage étant de type appliqué, nous parlerons de certains problèmes que l'on rencontre et de certaines méthodes que l'on emploie en pratique. Notre discussion s'applique

1. J.N.K. Rao, professeur et chercheur distingué, Université Carleton, Ottawa, Canada, K15 5B6. Courriel : jrao@math.carleton.ca; Wayne A. Fuller, professeur distingué émérite, Iowa State University, Ames, IA, É.-U. 50011. Courriel : waf@iastate.edu.

surtout aux grands échantillons polyvalents, c'est-à-dire les enquêtes au sujet desquelles nous avons le plus d'expérience. De même, notre connaissance des applications est concentrée au Canada et aux États-Unis.

La présentation de l'article est la suivante. La section 2 décrit les premières contributions marquantes de 1920 à 1960. Les problèmes inférentiels sont traités à la section 3. L'article se termine par une discussion sur l'avenir à la section 4.

2 Premières contributions marquantes : 1920 à 1960

Kiaer (1897) fut sans doute le premier à promouvoir l'échantillonnage (ou ce que l'on appelait à l'époque la méthode représentative) au lieu du dénombrement complet (recensement), quoique la référence la plus ancienne remonte à l'an 1000 avant notre ère. L'objectif de la méthode représentative est d'obtenir un échantillon qui reflète la population finie parente, ce que l'on peut réaliser par échantillonnage équilibré sur des totaux auxiliaires connus, par échantillonnage par choix raisonné ou par échantillonnage aléatoire menant à des probabilités d'inclusion égales. Dès les années 1920, la méthode représentative était d'usage très répandu. L'Institut international de statistique (IIS) a joué un rôle essentiel en créant un comité d'experts chargé de faire rapport sur cette méthode. La contribution de Bowley (1926) au rapport de l'IIS comprend ses travaux fondamentaux sur l'échantillonnage aléatoire stratifié avec répartition proportionnelle menant à des probabilités d'inclusion égales. Bowley (1936) déclare que la [*Traduction*] « première application de ce principe » d'inférer pour la population à partir d'un échantillon a été l'étude dans *Reading* en 1912. Bowley a recommandé que la méthode d'échantillonnage pour cette étude soit un échantillonnage systématique d'une liste de maisons. Bowley a appelé cette procédure systématique une [*Traduction*] « méthode pure d'échantillonnage » et a déclaré [*Traduction*] « C'est littéralement la méthode d'échantillonnage stratifié ». Bowley donne des exemples où l'échantillonnage systématique a été utilisé après 1912. Bowley (1936) souligne l'importance d'une base complète et des probabilités de sélection égales. Mais c'est Neyman (1934) qui a jeté les bases de l'échantillonnage probabiliste (ou approche fondée sur le plan de sondage). Il a démontré que l'échantillonnage aléatoire stratifié était préférable à l'échantillonnage équilibré (représentatif) tel qu'il était utilisé à l'époque. Il a également présenté le concept d'efficacité et de répartition optimale de l'échantillon, appelée aujourd'hui répartition de Neyman, qui minimise la taille totale de l'échantillon pour une précision spécifiée en relâchant la condition de probabilités d'inclusion égales de Bowley. En fait, Tchuprow (1923) avait obtenu la répartition de Neyman 10 ans plus tôt, dans un article découvert après la parution de celui de Neyman. Neyman (1934) a également montré que, pour des échantillons suffisamment grands, on pouvait obtenir pour la moyenne de population d'une variable d'intérêt des intervalles de confiance tels que la fréquence des erreurs mentionnée dans l'énoncé de confiance sous échantillonnage répété ne dépasse pas la limite établie, « quelles que soient les propriétés inconnues de la population ». Plus récemment, l'échantillonnage équilibré, préconisé au départ par Gini et Galvani, a été peaufiné de manière à intégrer les caractéristiques intéressantes de l'échantillonnage probabiliste ainsi que de l'échantillonnage équilibré sur des totaux auxiliaires connus (Deville et Tillé, 2004). La nouvelle méthode d'échantillonnage équilibré est maintenant appliquée en Europe, surtout en France, pour tirer des échantillons pour les enquêtes auprès des établissements. Une deuxième méthode de

sélection probabiliste contrôlée est l'échantillonnage réjectif, introduit par Hájek (1964) comme méthode de contrôle de la taille de l'échantillon dans l'échantillonnage de Poisson. Fuller (2009a) a étendu la procédure de manière à restreindre les échantillons acceptables à l'ensemble pour lequel les estimations de la moyenne des variables auxiliaires sont proches de la moyenne de population.

Au cours des années 1930, alors que s'accélérait la demande de renseignements socioéconomiques, les avantages de l'échantillonnage probabiliste, dont la réalisation d'études de plus grande portée, à moindre coût et plus rapidement que des recensements, ont bientôt été reconnus partout dans le monde. On a donc assisté à une multiplication et à une diversification des enquêtes fondées sur l'échantillonnage probabiliste et couvrant de grandes populations. Accepté presque universellement, l'échantillonnage probabiliste (ou approche fondée sur le plan de sondage) de Neyman est devenu un outil standard pour la recherche empirique en sciences sociales et en statistique officielle. On a également reconnu que la précision d'un estimateur dépendait en grande partie de la taille de l'échantillon et non de la fraction d'échantillonnage. Les années 1940 ont vu des études portant sur les propriétés de l'échantillonnage systématique pour différentes populations, voir Madow et Madow (1944), Cochran (1946), et Yates (1948). Cochran (1977, chapitre 8) présente une excellente discussion sur l'échantillonnage systématique clarifiant pourquoi seuls les estimateurs de variance fondés sur le modèle sont possibles. Voir aussi Bellhouse (1988). Au début, l'élaboration de la théorie de l'échantillonnage était axée sur l'estimation des totaux et des moyennes, et des erreurs d'échantillonnage connexes. Les erreurs non dues à l'échantillonnage, dont la non-réponse, les erreurs de couverture et les erreurs de mesure, étaient en grande partie ignorées en recherche théorique.

Passons maintenant à quelques avancées théoriques importantes concernant l'approche fondée sur le plan de sondage réalisées après Neyman. Dès 1937, Mahalanobis a utilisé des plans d'échantillonnage à plusieurs degrés pour les enquêtes sur les récoltes en Inde. Dans son article classique publié en 1944 (Mahalanobis, 1944), il a formulé rigoureusement les fonctions de coût et de variance permettant une conception efficace des enquêtes. Il a joué un rôle essentiel dans la création du *National Sample Survey* de l'Inde, la plus grande enquête polyvalente continue effectuée par du personnel à temps plein réalisant des interviews sur place pour les enquêtes socioéconomiques et des mesures physiques pour les enquêtes sur les récoltes. Sukhatme, qui fut l'élève de Neyman, a également contribué de manière novatrice à la conception et à l'analyse des enquêtes agricoles à grande échelle en Inde, en faisant appel à l'échantillonnage à plusieurs degrés stratifié. Les traités classiques sur l'échantillonnage publiés par Cochran (1953), Deming (1950), Hansen, Hurwitz et Madow (1953), Sukhatme (1954) et Yates (1949) ont été utiles aux étudiants ainsi qu'aux praticiens.

Durant la période de 1940 à 1960, sous la direction de Morris Hansen, les statisticiens d'enquête du *U.S. Census Bureau* ont contribué de manière fondamentale à la théorie et à la méthodologie des enquêtes par sondage. Cette période est considérée comme l'âge d'or du *Census Bureau*. Hansen et Hurwitz (1943) ont élaboré la théorie de base de l'échantillonnage en grappes à deux degrés stratifié, avec tirage d'une grappe (ou unité primaire d'échantillonnage) dans chaque strate avec probabilité proportionnelle à la taille (PPT), puis sous-échantillonnage de ces grappes à un taux assurant un échantillon autopondéré (probabilités globales de sélection égales). La sélection de grappes avec probabilités inégales peut réduire considérablement la variance en contrôlant la variabilité découlant des tailles de grappe inégales. Une autre contribution importante du *U.S. Census Bureau* a été l'utilisation de l'échantillonnage rotatif avec

renouvellement partiel des ménages pour résoudre le problème du fardeau de réponse dans les enquêtes répétées au cours du temps, comme la *Current Population Survey* réalisée mensuellement aux États-Unis pour mesurer les taux de chômage. Hansen, Hurwitz, Nisselson et Steinberg (1955) ont établi des estimateurs composites simples, mais efficaces sous échantillonnage rotatif. L'usage de l'échantillonnage rotatif et de l'estimation composite est très répandu dans les enquêtes continues à grande échelle.

Avant les années 1950, l'objectif principal était l'estimation des totaux et des moyennes de population. Au *U.S. Census Bureau*, Woodruff (1952) a élaboré une approche unifiée de construction d'intervalles de confiance pour les quantiles (en particulier, la médiane) applicable aux plans d'échantillonnage généraux. La procédure demeure l'une des pierres angulaires de l'estimation des quantiles (Francisco et Fuller, 1991).

Après la consolidation de la théorie fondamentale de l'échantillonnage fondé sur le plan de sondage, Hansen, Hurwitz, Marks et Mauldin (1951) et d'autres se sont penchés sur les erreurs de mesure ou de réponse dans les données d'enquêtes. Sous des modèles d'erreur de mesure additifs avec hypothèses de modèle minimales sur les réponses observées traitées comme des variables aléatoires, la variance totale d'un estimateur peut être décomposée en une variance d'échantillonnage, une variance de réponse simple et une variance de réponse corrélée (VRC) attribuable aux intervieweurs.

Mahalanobis (1946) a établi la méthode des sous-échantillons interpénétrants pour évaluer les erreurs d'échantillonnage ainsi que d'intervieweur. En répartissant les sous-échantillons au hasard entre les intervieweurs, il est possible d'estimer la variance totale ainsi que la composante due à l'intervieweur. Cette dernière peut dominer la variance totale quand le nombre d'intervieweurs est faible. En vue d'éliminer la composante VRC due aux intervieweurs, un autodénombrement par la poste a été lancé pour le Recensement des États-Unis de 1960.

La question de la non-réponse aux enquêtes a également été abordée durant les premiers travaux sur l'échantillonnage. Hansen et Hurwitz (1946) ont proposé l'échantillonnage à deux phases qui consiste à prendre contact par la poste avec l'échantillon sélectionné à la première phase, puis à procéder à l'interview sur place d'un sous-échantillon de non-répondants, en supposant que la réponse sera complète ou que la non-réponse sera négligeable à la deuxième phase. Cette méthode a été utilisée récemment au Canada quand l'échantillon pour le questionnaire long à réponse obligatoire du recensement a été remplacé par l'Enquête nationale auprès des ménages à participation volontaire. Après le changement de gouvernement en 2015, le premier ministre du Canada a rétabli le questionnaire long du Recensement. L'échantillonnage à deux phases continuera d'être utilisé, mais dans une moindre mesure. La méthode d'échantillonnage à deux phases de Hansen-Hurwitz a aussi été utilisée dans d'autres enquêtes incluant l'*American Community Survey*.

Une attention particulière a aussi été accordée aux inférences pour des sous-populations non planifiées (appelées domaines), telles que les groupes âge-sexe à l'intérieur d'un État. Hartley (1959) et Durbin (1958) ont élaboré une théorie unifiée pour l'estimation de domaines applicable à des plans de sondage généraux et ne nécessitant que des formules existantes pour le calcul des totaux et des moyennes de population.

Durant cette première période, la théorie de l'échantillonnage a été établie en majeure partie par des praticiens de la statistique officielle, tandis que les chercheurs universitaires, surtout aux États-Unis, accordaient peu d'attention à l'échantillonnage. Faisait exception la *Iowa State University*, où le corps

professoral a joué un rôle de chef de file dès le départ sous la direction de Cochran, Jessen et Hartley. Une autre institution qui a contribué très tôt à la pratique des enquêtes et à la recherche sur ces derniers est le *Survey Research Center* de l'Université du Michigan créé en 1947 et dont Leslie Kish fut un des premiers membres.

Au cours des années 1950, des cadres théoriques formels pour l'inférence fondée sur le plan de sondage pour des totaux et des moyennes ont été proposés en considérant les données d'échantillon comme un ensemble d'étiquettes d'échantillon groupé avec les variables d'intérêt associées. Horvitz et Thompson (1952) ont dérivé l'estimateur bien connu dans lequel la pondération est inversement proportionnelle à la probabilité d'inclusion. Narain (1951) a également proposé cet estimateur. Godambe (1955) a élaboré une classe générale d'estimateurs linéaires en permettant que le poids de sondage d'une unité dépende de l'étiquette de celle-ci, ainsi que des étiquettes des autres unités dans l'échantillon. Il a ensuite montré que le meilleur estimateur linéaire sans biais n'existe pas dans cette classe générale, même sous échantillonnage aléatoire simple.

3 Problèmes d'inférence : 1950 -

3.1 Fondements théoriques

Des tentatives ont eu lieu en vue d'intégrer la théorie des enquêtes par sondage à l'inférence statistique traditionnelle en se servant de la fonction de vraisemblance. Godambe (1966) a montré que la fonction de vraisemblance issue de l'ensemble complet de données d'échantillon, y compris les étiquettes, en considérant le vecteur des valeurs de population inconnues comme étant le paramètre, ne fournit aucun renseignement sur les valeurs non échantillonnées ni, donc, sur le total ou la moyenne de population. Cet aspect non informatif de la fonction de vraisemblance est dû à l'inclusion des étiquettes dans l'ensemble de données, ce qui rend l'échantillon unique. Une autre route fondée sur le plan de sondage ne tient pas compte de certains aspects des données d'échantillon afin de rendre l'échantillon non unique et, donc, arriver à des fonctions de vraisemblance informatives (Hartley et Rao, 1968; Royall, 1968). Cette approche de la vraisemblance non paramétrique est similaire à l'approche de la vraisemblance empirique (VE) populaire à l'heure actuelle en inférence statistique classique (Owen, 1988). L'approche de la vraisemblance empirique a été appliquée récemment aux problèmes d'échantillonnage pour estimer non seulement les totaux et les moyennes, mais aussi des paramètres plus complexes. Ainsi, les efforts d'intégration à la statistique classique ont été partiellement fructueux.

L'approche dépendante d'un modèle offre un autre chemin vers l'inférence à partir de données d'enquêtes. Cette approche requiert que la structure de la population obéisse à un modèle de superpopulation spécifié. La distribution induite par le modèle supposé sert de base pour les inférences (Brewer, 1963 et Royall, 1970). Ce genre d'inférences conditionnelles (à l'échantillon) peut être intéressant. Cependant, les estimateurs résultants sont parfois non convergents par rapport au plan et, par conséquent, peuvent donner de mauvais résultats pour les grands échantillons si la spécification du modèle est incorrecte (Hansen, Madow et Tepping, 1983).

Une approche hybride, appelée approche assistée par modèle, vise à combiner les caractéristiques désirables des méthodes fondées sur le plan de sondage et de celles dépendantes d'un modèle; consulter Cassel, Särndal et Wretman (1976). Cette approche comprend habituellement l'utilisation de données ne provenant pas de la collecte des données, appelées données auxiliaires. Les procédures qui font appel à des données auxiliaires comprennent l'estimation par la régression, l'estimation par le ratio et le *raking* (ou méthode de ratissage), c'est-à-dire des méthodes avec estimateurs linéaires en la variable d'intérêt. La force des estimateurs utilisant de l'information auxiliaire, particulièrement les estimateurs par régression, a été reconnue très tôt (Cochran, 1953). Au cours des années 1970, grâce à la puissance des ordinateurs, l'estimation par la régression est devenue réalisable, mais pour être acceptable dans les enquêtes à grande échelle, il faut que les poids de régression ne soient pas négatifs. Une première définition des poids non négatifs figure dans Huang et Fuller (1978). Deville et Särndal (1992) ont donné une méthode générale de construction des poids pour des estimateurs convergents par rapport au plan. Les méthodes assistées par modèle ne concernent que les estimateurs du total convergents par rapport au plan qui sont également sans biais sous le modèle pour un modèle de travail. Cette approche est utile pour les grands échantillons et aboutit à des inférences fondées sur le plan de sondage acceptables, indépendamment de la validité du modèle de travail. Cependant, l'efficacité des estimateurs dépend du degré auquel le modèle de travail s'approche de la vraie structure de population. La catégorie la plus répandue d'estimateurs assistés par modèle est celle des estimateurs par la régression généralisée (GREG) qui sont mis en œuvre dans les progiciels pour données d'enquêtes.

Les résultats théoriques concernant l'échantillonnage probabiliste mettent l'accent sur les deux premiers moments de la statistique d'échantillon. Des théorèmes centraux limites ont été utilisés pour justifier des intervalles de confiance fondés sur la normalité. Un des premiers théorèmes centraux limites pour des échantillons aléatoires simples est celui de Madow (1948). Hájek (1960) a donné un théorème central limite pour l'échantillonnage aléatoire simple et un théorème pour l'échantillonnage réjectif dans Hájek (1964). Bickel et Freedman (1984) ont donné un théorème central limite pour l'échantillonnage aléatoire stratifié. Une publication récente prend en considération des séries de populations finies fixes ainsi que des séries de populations finies qui sont des échantillons tirés d'une superpopulation (Fuller, 2009b, section 1.3.2).

Durant les années 1930 et les années 1940, le coût de l'estimation de la variance était très élevé, presque prohibitif, et il demeure élevé aujourd'hui. Dès le départ, le rééchantillonnage a été adopté comme méthode efficace d'estimation de la variance. Ainsi que nous l'avons mentionné, Mahalanobis (1939, 1946) a proposé une forme précoce de rééchantillonnage, qu'il a appelé échantillons « interpénétrants » et que des auteurs ultérieurs ont appelé « groupes aléatoires ». La méthode des groupes aléatoires fondés sur des demi-échantillons a été utilisée par le *U.S. Census Bureau* dans les années 1950 et 1960. McCarthy (1966, 1969) a développé et décrit l'estimation de la variance pour les demi-échantillons équilibrés. Voir aussi Kish et Frankel (1970). Wolter (2007) présente une discussion élaborée sur les demi-échantillons équilibrés. Voir aussi Dippo, Fay et Morgenstein (1984), Kish et Frankel (1974), Krewski et Rao (1981), et Rao et Shao (1999). Le jackknife et le bootstrap sont les versions courantes des premières procédures de rééchantillonnage. Wolter (2007, chapitre 4) attribue à Durbin (1959) la première utilisation du jackknife dans l'estimation sur population finie. L'utilisation du bootstrap dans les conditions classiques remonte à Efron (1979), mais son application aux échantillons tirés avec probabilités inégales et aux populations finies

n'a pas été immédiate. Parmi le grand nombre d'articles traitant du jackknife et du bootstrap pour les échantillons d'enquête figurent ceux de McCarthy et Snowden (1985), qui présentent une première version de l'échantillonnage avec remise, et de Rao et Wu (1988), qui décrivent un bootstrap modifié fondé sur un « rééchelonnement » (*rescaling*) des poids pour les échantillons d'enquête. Sitter (1992) a abordé plusieurs questions et, notamment, a fait des suggestions pour obtenir des tailles d'échantillon entières. Antal et Tillé (2011) ont exposé des méthodes bootstrap appropriées pour une grande gamme de plans de sondage, y compris l'échantillonnage de Poisson. Beaumont et Patak (2012) ont proposé des procédures bootstrap générales.

3.2 Usage analytique des données d'enquêtes

Comme nous l'avons souligné, les premiers travaux sur l'échantillonnage probabiliste étaient axés sur les totaux et les moyennes, et bon nombre de procédures d'estimation ont été élaborées pour la statistique officielle. Cependant, dès le début, les spécialistes des sciences sociales ont utilisé des échantillons d'enquête pour obtenir des réponses aux questions dans ce domaine applicables au-delà de la population finie échantillonnée. Deming et Stephan (1940) et Deming (1953) ont tenu compte explicitement de la différence entre les usages « énumératif » et « analytique » des données d'enquête et de recensement; voir aussi Hartley (1959). Les estimations analytiques sont parfois appelées estimations pour une superpopulation. Les premiers analystes traitaient souvent les données d'enquête par sondage comme provenant d'un échantillon aléatoire simple et construisaient des estimations sur cette base. Puisque ne pas tenir compte du plan de sondage pouvait introduire un biais, une théorie de l'estimation a été élaborée pour les estimations analytiques. L'une de ses composantes consistait en des tests pour évaluer les effets des poids sur les estimations; voir DuMouchel et Duncan (1983), Fuller (1984), et Korn et Graubard (1995). Une deuxième composante était l'élaboration d'une théorie fondée sur le plan de sondage pour les statistiques compliquées. Voir Fuller (1975), Rao et Scott (1981, 1984), ainsi que Binder et Roberts (2003). La troisième composante consistait à intégrer le plan de sondage dans le modèle (Skinner, 1994 et Pfeffermann et Sverchkov, 1999). Il existe aujourd'hui un certain nombre de progiciels (SAS, SUDAAN, R, STATA) pour calculer les statistiques et les erreurs-types fondées sur l'approche probabiliste. Nombre des algorithmes remontent aux travaux effectués à la *Iowa State University* (Hidiroglou, Fuller et Hickman, 1976).

3.3 Données manquantes

Presque tous les échantillons (et toutes les expériences) présentent des données manquantes ou incorrectes. Dans le cas des enquêtes par sondage, les données manquantes sont réparties en deux catégories, à savoir les unités manquantes (non-réponse totale) et les questions manquantes (non-réponse partielle), l'expression unité manquante signifiant que tous les items (variables) manquent dans l'enregistrement de la réponse de l'unité. L'ensemble de monographies publié par Madow, Nisselson et Olkin (1983) témoigne de l'importance des données manquantes dans les études par sondage. L'une des méthodes de traitement des données manquantes se résumait à communiquer la nature et le nombre des items manquants et à totaliser les items restants. Au début, cette méthode était courante, mais l'hypothèse implicite d'interchangeabilité

sur laquelle elle reposait n'était pas souvent raisonnable. Une des premières méthodes de correction de la non-réponse totale consistait à utiliser un répondant de substitution, ce qui revenait souvent à interviewer une personne « proche » du non-répondant. Une modification courante à l'étape de l'analyse était, et demeure, la poststratification (Deming, 1953; Thomsen, 1973; Kalton, 1983 et Jagers, 1986). Dans la littérature sur les données manquantes, les poststrates sont souvent appelées cellules. Les estimateurs par la régression, qui sont des extensions directes des estimateurs de cellule, représentent une méthode importante de correction des données manquantes (Fuller et An, 1998). Les méthodes de pondération en vue de traiter la non-réponse totale sont passées en revue par Brick et Montaquila (2009).

Diverses formes d'imputation pour corriger la non-réponse partielle ont été appliquées au fil du temps, l'imputation étant effectuée manuellement avant l'usage des ordinateurs. L'une des premières imputations formelles fondées sur un modèle et effectuées par ordinateur était la procédure d'imputation hot deck utilisée par le *U.S. Census Bureau* pour la *Current Population Survey* de 1947; voir la description dans Andridge et Little (2009). L'accroissement de la puissance de calcul et les progrès techniques (Little, 1982; Kalton et Kish, 1984; Rubin, 1974, 1976, 1987; Little et Rubin, 1987; Kim et Fuller, 2004) ont fait de l'imputation un élément standard de l'estimation sur des échantillons d'enquête et un domaine de recherche dynamique. Parmi les ouvrages récents à ce sujet, mentionnons Kim et Shao (2013) et Little et Rubin (2014).

3.4 Estimation sur petits domaines

L'usage croissant de modèles pour l'estimation sur petits domaines résulte de la conjonction de deux facteurs. Le premier est la demande d'estimations pour de petits domaines (par exemple, régions géographiques) pour la formulation de politiques, la répartition des fonds et la planification régionale. Le second est la grande erreur-type associée à de nombreux estimateurs de domaines sous le plan de sondage. Schaible (1996) et Purcell et Kish (1979) ont donné tôt des exemples d'estimation sur petits domaines; voir aussi Gonzalez (1973) et Steinberg (1979). Dès 1947, le *U.S. Census Bureau* utilisait des méthodes fondées sur un modèle pour l'estimation sur petits domaines (Hansen et coll., 1953; Vol. I, pages 483-486). Plus récemment, les modèles linéaires mixtes incluant des effets fixes et aléatoires ont pris de l'importance. Les premiers usages de modèles mixtes pour l'estimation sur petits domaines sont décrits dans Fay et Herriot (1979) et dans Battese, Harter et Fuller (1988). Certains ensembles d'estimations sur petits domaines peuvent être considérés comme une réaffectation des estimations de domaines, en maintenant l'estimation directe du total général convergente par rapport au plan. Les méthodes bayésiennes, en particulier les méthodes bayésiennes hiérarchiques, sont de plus en plus fréquemment utilisées, en raison de la capacité à traiter des modèles complexes; voir Rao et Molina (2015, chapitre 10). En raison de la demande croissante, les publications en la matière se sont multipliées et l'estimation sur petits domaines fait aujourd'hui l'objet d'assemblées de spécialistes fréquentes, ainsi que d'un livre (Rao, 2003) dont la seconde édition est parue récemment (Rao et Molina, 2015).

3.5 Pratiques des enquêtes

Les questions relatives aux plans de sondage et à l'estimation dont nous venons de discuter sont des aspects essentiels d'une opération d'enquête, mais ne représentent qu'une petite fraction de l'ensemble. La

qualité du produit fini dépend des données qui figurent dans la base de sondage, de l'instrument de collecte, de la collecte, de la vérification et du traitement des données, ainsi que de la présentation des résultats. De nombreuses sources d'erreur sont difficiles à mesurer, mais les concepteurs des enquêtes font des estimations de coût implicites quand ils répartissent les ressources entre les différents volets de l'opération d'enquête. Groves et Lyberg (2010) passent en revue les tentatives en vue d'énumérer les composantes de la qualité d'une enquête et de les regrouper sous un chapiteau unique. Ils attribuent à Deming (1944) une première description des sources d'erreur dans les enquêtes par sondage et décrivent les contributions de Dalenius (1974), Anderson, Kasper et Frankel (1979), Groves (1989), Biemer et Lyerg (2003), entre autres. Groves et Herringa (2006) ont proposé des outils de contrôle dynamique des erreurs et des coûts d'enquête qui permettent d'aboutir à des plans de collecte adaptatifs pour les enquêtes auprès des ménages. En particulier, les paradonnées (mesures concernant le processus de collecte des données d'enquêtes) peuvent servir à surveiller le travail sur le terrain, à décider quand il faut intervenir durant la collecte des données, et à traiter l'erreur de mesure, la non-réponse et les erreurs de couverture (Kreuter, 2013).

4 L'avenir

Nous pouvons projeter un certain nombre de situations courantes dans l'avenir. Le contexte budgétaire sera strict et les demandes de produits augmenteront. Il y aura une demande de prévisions, ainsi que d'amélioration de l'accès des utilisateurs aux données. Il y aura une demande d'accélération de la production des statistiques, et ce, naturellement, sans compromettre la qualité. Une pression sera exercée afin que concordent les estimations provenant de sources différentes.

Nous nous attendons à ce que la plus grande vitesse des ordinateurs influe sur tous les aspects du domaine. Des algorithmes de vérification et d'imputation plus complexes seront établis. Le temps écoulé entre la collecte et la publication sera raccourci. Des analyses plus complexes seront effectuées sur les données d'enquêtes. Les procédures de couplage d'enregistrements seront améliorées. Les données seront mises à la disposition des utilisateurs sous différentes formes. Les bases de données consultables permettant à l'utilisateur de faire des recherches deviendront plus courantes. L'utilisation de données auxiliaires de toutes sortes, et en particulier des données administratives, augmentera. Les données administratives seront utilisées à la fois comme données auxiliaires et comme estimations directes pour certaines variables. Citro (2014) donne des exemples de variables pour lesquels des données administratives peuvent remplacer les réponses aux questions dans un questionnaire. L'utilisation de données auxiliaires quand l'appariement aux données recueillies est imparfait deviendra un domaine de recherche.

Les méthodes de communication modernes et les médias sociaux sont à l'origine de vastes quantités de données, la plupart d'entre elles générées pour répondre à des objectifs de court terme et mal spécifiés. Le terme « mégadonnées » n'est pas bien défini, mais la plupart d'entre nous seraient d'accord que les données des médias sociaux en font partie. Le rapport de l'AAPOR sur les mégadonnées (2015) contient une excellente analyse des possibilités et des défis associés à ce type de données. Tam et Clarke (2015) et Pfeffermann (2015) discutent de ces questions dans la perspective d'un organisme statistique gouvernemental. Parce qu'ils font partie de la société moderne, les médias sociaux présentent en eux-mêmes un intérêt pour les spécialistes des sciences sociales. Donc, des index et des résumés de ces données sont, et

seront, produits. À titre d'exemple, mentionnons le *Social Media Job Loss Index* de l'Université du Michigan. L'échantillonnage joue un grand rôle dans la création de produits à partir de ces données.

L'un des défis consiste à transformer certains types de mégadonnées en une forme utilisable comme données auxiliaires. Par exemple, Porter, Holan, Wikle et Cressie (2014) se sont servis des fréquences de recherche de mots espagnols fournies par l'outil Google Trends comme covariables fonctionnelles pour estimer, au niveau de l'État, les proportions de personnes parlant l'espagnol, en utilisant les estimations de l'*American Community Survey* comme variables dépendantes dans des modèles de petits domaines.

L'un des avantages souvent mentionnés des enquêtes par sondage comparativement aux recensements est le coût. La structure de coût a évolué avec l'accroissement de la puissance de calcul et semble destinée à continuer d'évoluer. Aux États-Unis, la *National Land Cover Database* est un recensement de la couverture des terres (Han, Yang, Di et Mueller, 2012). En principe, les procédures de classification s'amélioreront, si bien que les données de ce genre serviront plus souvent de données auxiliaires. Les organismes de collecte de données investiront davantage dans la construction de fichiers de données auxiliaires améliorés au niveau de la population, de sorte que certaines données recueillies aujourd'hui sur des échantillons le seront au niveau de la population. Les mêmes types de développement des données se poursuivront pour les statistiques sur les populations et les entreprises.

Nécessairement, notre discussion s'étend peu sur la collecte des données. La façon dont l'évolution de la technologie a modifié les procédures de collecte des données est peut-être plus évidente que le lien entre la technologie et la théorie. Pour les liens avec la théorie, consulter Bellhouse (2000). La collecte des données assistée par ordinateur devient la norme. Il faut s'attendre à ce que l'utilisation de la technologie de géolocalisation s'accroisse. Il est raisonnable de prévoir un plus grand usage d'outils de télédétection et de collecte des données à distance. Par exemple, il serait facile d'intégrer des données physiques recueillies par un appareil comme la montre Apple ou Fitbit dans une étude sur la santé. Des appareils de surveillance plus gros et moins attrayants sont utilisés à l'heure actuelle dans les enquêtes sur l'activité physique (van Remoortel, Giavedoni, Raste, Burtin, Louvaris, Gimeno-Santos, Langer, Glendenning, Hopkinson, Vogiatzis, Peterson, Wilson, Mann, Rabinovich, Puhan, Troosters et PROactive consortium, 2012).

Comme le révèle l'expérience récente, il devient de plus en plus difficile de recueillir des données par interview téléphonique ou par interview sur place. Les enquêtés font face à des activités plus vastes de collecte systématique de données. L'omniprésence des questionnaires sur la satisfaction concernant toute chose, des services médicaux au dentifrice, doit indubitablement avoir une incidence sur la volonté des individus à répondre. Il semble raisonnable de prévoir qu'il sera de plus en plus difficile d'obtenir leur coopération pour les méthodes habituelles de collecte des données. Cette tendance entraînera une étude plus approfondie de la nature des non-répondants et de la non-réponse. De même, des efforts seront déployés en vue d'adapter la collecte des données à l'évolution des méthodes de communication.

Les échantillons non probabilistes ont fait partie de l'activité de sondage tout au long de la période ultérieure à Neyman. En particulier, l'échantillonnage par quotas est couramment utilisé dans la recherche en marketing et ailleurs pour des raisons de coût (Sudman, 1966; 1976). Moser et Stuart (1953) et Stephan et McCarthy (1958) ont fait les premières comparaisons entre l'échantillonnage par quotas et l'échantillonnage probabiliste. Cochran (1977, page 136) affirme [*Traduction*] « La méthode des quotas

produira probablement des échantillons biaisés pour les caractéristiques telles que le revenu, l'éducation et l'occupation, même si elle s'accorde souvent avec les échantillons probabilistes pour les questions d'opinion et d'attitude ». Le recours à des procédures telles que la poststratification et l'estimation par la régression sur des échantillons non probabilistes s'est poursuivi au même rythme que l'utilisation sur des échantillons probabilistes. Le changement de nature de la communication humaine offre des possibilités en ce qui concerne tant les procédures fondées sur un modèle que celles fondées sur les probabilités. En raison des structures de coût, de nouvelles méthodes, telles les procédures axées sur le Web, seront souvent appliquées en premier lieu dans des conditions non probabilistes et à des fins non gouvernementales.

À mesure que les procédures d'appariement s'amélioreront et que la demande de données détaillées augmentera, les procédures de contrôle de la divulgation et les travaux de recherche connexes susciteront davantage d'attention.

L'échantillonnage dans le contexte des enquêtes est une discipline appliquée, dont le fonctionnement est défini par la conjoncture sociale, géographique, culturelle et technologique. Il est difficile de prévoir l'impact qu'auront les changements sociaux et culturels sur cette discipline, même à court terme. Le fait que l'on doive présumer aujourd'hui que presque toutes nos activités publiques et une grande partie de nos activités privées sont susceptibles d'être enregistrées entraînera-t-il une attitude plus détendue de réponse aux questions ? L'amélioration des outils de surveillance fera-t-elle que les enquêtés seront davantage disposés à ce que l'on surveille leurs activités physiques ? Ou bien toute cette surveillance secondaire entraînera-t-elle une réaction négative à l'égard de la collecte systématique de données ? La plus grande disponibilité de résultats fondés sur des données que l'on a recueillies aura-t-elle un effet positif ou négatif sur les efforts de collecte de données ? Quel sera l'impact des divers médias sociaux ?

Il ressort clairement de la présente discussion que des facteurs externes à notre discipline détermineront nos futures activités. Il nous faudra alors adapter la collecte, le traitement, ainsi que la présentation et la diffusion des données.

Remerciements

Nous remercions Graham Kalton pour ses remarques et suggestions ayant mené à des améliorations dans la version originale. Nous remercions les quatre participants, Graham Kalton, Sharon Lohr, Danny Pfeffermann et Chris Skinner, pour les suppléments sur l'histoire, les observations perspicaces sur le présent et les commentaires sur le futur de l'échantillonnage. Nous avonc choisi de ne pas préparer de réponse à ces discussions, car nous en avons apprécié une grande partie et avons trouvé peu de fondement à un désaccord.

Bibliographie

AAPOR Big Data Task Force (2015). *AAPOR Report on Big Data*. https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/images/BigDataTaskForceReport_FINAL_2_12_15_b.pdf.

Anderson, R., Kasper, J. et Frankel, M. (1979). *Total Survey Error: Applications to Improve Health Surveys*. San Francisco, CA: Jossey-Bass.

Andridge, R.H., et Little, R.J. (2009). The use of sample weights in hot deck imputation. *Journal of Official Statistics*, 25, 21-36.

Antal, E., et Tillé, Y. (2011). A direct bootstrap method for complex sampling designs from a finite population. *Journal of the American Statistical Association*, 106, 534-543.

Battese, G.E., Harter, R.M. et Fuller, W.A. (1988). An error component model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83, 28-36.

Beaumont, J.-F., et Patak, Z. (2012). On the generalized bootstrap for sample surveys with special attention to Poisson sampling. *International Statistical Review*/*Revue Internationale de Statistique*, 80, 127-148.

Bellhouse, D.R. (1988). Systematic sampling. Dans *Handbook of Statistics*, (Éds., P.R. Kreshnaiah et C.R. Rao), Elsevier, 6, 125-145.

Bellhouse, D.R. (2000). Évolution de la théorie de l'échantillonnage d'enquête au XX^{e} siècle et son rapport avec l'informatique. *Techniques d'enquête*, 26, 1, 13-23. Article accessible à l'adresse http://www.statcan.gc.ca/pub/12-001-x/2000001/article/5174-fra.pdf..

Bethlehem, J. (2009). The rise of survey sampling. Discussion Paper (09015), Statistics Netherlands, La Haye.

Bickel, P.J., et Freedman, D.A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *The Annals of Statistics*, 12, 470-482.

Biemer, P.P., et Lyberg, L. (2003). *Introduction to Survey Quality*. New York: John Wiley & Sons, Inc.

Binder, D.A., et Roberts, G.A. (2003). Design-based and model-based methods for estimating model parameters. Dans *Analysis of Survey Data*, (Éds., R.L. Chambers et C.J. Skinner), Wiley, Chichester, Royaume-Uni, 29-48.

Bowley, A.L. (1926). Measurement of the precision attained in sampling. *Bulletin de l'Institut International de Statistique*, 22(1), 1-62.

Bowley, A.L. (1936). The application of sampling to economics and sociological problems. *Journal of the American Statistical Association*, 31, 474-480.

Brewer, K.R.W. (1963). Ratio estimation and finite populations: Some results deducible from the assumption of an underlying stochastic process. *Australian Journal of Statistics*, 5, 93-105.

Brewer, K.R.W. (2013). Trois controverses dans l'histoire de l'échantillonnage. *Techniques d'enquête*, 39, 2, 275-289. Article accessible à l'adresse http://www.statcan.gc.ca/pub/12-001-x/2013002/article/11883-fra.pdf.

Brick, M.J. (2011). The future of survey sampling. *Public Opinion Quarterly*, 75, 872-888.

Brick, M.J., et Montaquila, J.M. (2009). Non response and weights. Dans *Handbook of Statistics*, (Éds., D. Pfeffermann et C.R. Rao), Elsevier, Amsterdam, 29A, 163-185.

Cassel, C.M., Särndal, C.-E. et Wretman, J.H. (1976). Some results on generalized difference estimation and generalized regression estimation for finite populations. *Biometrika*, 63, 615-620.

Citro, C.F. (2014). Des modes multiples pour les enquêtes à des sources de données multiples pour les estimations. *Techniques d'enquête,* 40, 2, 151-181. Article accessible à l'adresse http://www.statcan.gc.ca/pub/12-001-x/2014002/article/14128-fra.pdf.

Cochran, W.G. (1953). *Sampling Techniques.* New York: John Wiley & Sons, Inc.

Cochran, W.G. (1977). *Sampling Techniques, 3rd Edition.* New York: John Wiley & Sons, Inc.

Dalenius, T. (1974). Ends and Means of Total Survey Design. Report in "Errors in Surveys", Stockholm University.

Deming, E. (1944). On errors in surveys. *American Sociological Review,* 9, 359-369.

Deming, E. (1950). *Some Theory of Sampling*. New York: John Wiley & Sons, Inc.

Deming, W.E. (1953). On a probability mechanism to attain an economic balance between the resultant error of non-response and the bias of non-response. *Journal of the American Statistical Association,* 48, 743-772.

Deming, W.E., et Stephan, F.F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics,* 11, 4, 427-444.

Deville, J.-C., et Särndal, C.-E. (1992). Calibration-estimators in survey sampling. *Journal of the American Statistical Association,* 376-382.

Deville, J.-C., et Tillé, Y. (2004). Efficient balanced sampling: The cube method. *Biometrika,* 91, 893-912.

Dippo, C.S., Fay, R.E. et Morgenstein, D.H. (1984). Computing variances from complex samples with replicate weights. *Proceedings of the American Statistical Association, Section on Survey Research Methods,* 489-494.

DuMouchel, W.H., et Duncan, G.J. (1983). Using sample survey weights in multiple regression analysis of stratified samples. *Journal of the American Statistical Association,* 78, 535-543.

Durbin, J. (1958). Sampling theory for estimates based on fewer individuals than the number selected. *Bulletin de l'Institut International de Statistique,* 36, 113-119.

Durbin, J. (1959). A note on the application of Quenouille's method of bias reduction to the estimation of ratios. *Biometrika,* 46, 477-480.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics,* 7, 1-26.

Fay, R.E., et Herriot, R.A. (1979). Estimates of income for small places. An application of James-Stein procedures to census data. *Journal of the American Statistical Association,* 74, 366, 269-277.

Francisco, C.A., et Fuller, W.A. (1991). Estimation of quantiles with survey data. *Annals of Statistics,* 19, 454-469.

Fuller, W.A. (1975). Regression analysis sample survey. *Sankhyā, Series C,* 37, 117-132.

Fuller, W.A. (1984). Application de la méthode des moindres carrés et de techniques connexes aux plans de sondage complexes. *Techniques d'enquête*, 10, 1, 107-130. Article accessible à l'adresse http://www.statcan.gc.ca/pub/12-001-x/1984001/article/14352-fra.pdf.

Fuller, W.A. (2009a). Some design properties of a rejective sampling procedure. *Biometrika*, 96, 1-12.

Fuller, W.A. (2009b). *Sampling Statistics*. New York: John Wiley & Sons, Inc.

Fuller, W.A., et An, A.B. (1998). Regression adjustments for nonresponse. *Journal of Indian Society of Agricultural Statistics*, 51, 331-342.

Godambe, V.P. (1955). A unified theory of sampling from finite populations. *Journal of the Royal Statistical Society, Series B*, 17, 269-278.

Godambe, V.P. (1966). A new approach to sampling from finite populations. *Journal of the Royal Statistical Society, Series B*, 28, 310-328.

Gonzalez, M.E. (1973). Use and evaluation of synthetic estimates. *Proceedings of the Social Statistics Section of the American Statistical Association*, 33-36.

Groves, R.M. (1989). *Survey Errors and Survey Costs.* New York: John Wiley & Sons, Inc.

Groves, R.M. (2011). Three eras of survey research. *Public Opinion Quarterly*, 73, 861-871.

Groves, R.M., et Heeringa, S.G. (2006). Responsive designs for household surveys: Tolls for actively controlling survey errors and costs. *Journal of the Royal Statistical Society, Series A*, 169, 439-457.

Groves, R.M., et Lyberg, L. (2010). Total survey error: Past, present and future. *Public Opinion Quarterly*, 74, 849-879.

Hájek, J. (1960). Limiting distributions in simple random sampling from a finite population. *Publications of the Mathematical Institute of the Hungarian Academy*, 5, 361-374.

Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Annals of Mathematical Statistics*, 35, 1491-1523.

Han, W., Yang, Z., Di, L. et Mueller, R. (2012). CropScape: A Web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support. *Computers and Electronics in Agriculture*, 84, 111-123.

Hansen, M.H., et Hurwitz, W.N. (1943). On the theory of sampling from finite populations. *Annals of Mathematical Statistics*, 14, 333-362.

Hansen, M.H., et Hurwitz, W.N. (1946). The problem of non-response in sample surveys. *Journal of the American Statistical Association*, 41, 517-529.

Hansen, M.H., Hurwitz, W.N. et Madow, W.G. (1953). *Sample Survey Methods and Theory*, Vols. I et II, New York: John Wiley & Sons, Inc.

Hansen, M.H., Madow, W.G. et Tepping, B.J. (1983). An evaluation of model-dependent and probability-sampling inferences in sample surveys. *Journal of the American Statistical Association*, 78, 776-793.

Hansen, M.H., Hurwitz, W.N., Marks, E.S. et Mauldin, W.P. (1951). Response errors in surveys. *Journal of the American Statistical Association,* 46, 147-190.

Hansen, M.H., Hurwitz, W.N., Nisselson, H. et Steinberg, J. (1955). The redesign of the census current population survey. *Journal of the American Statistical Association,* 50, 701-719.

Hartley, H.O. (1959). Analytical studies of survey data. Dans le volume en l'honneur de Corrado Gini, Instituto di Statistica, Rome, 1-32.

Hartley, H.O., et Rao, J.N.K. (1968). A new estimation theory for sample surveys. *Biometrika,* 55, 547-557.

Hidiroglou, M.A., Fuller, W.A. et Hickman, R.D. (1976). *SUPER CARP* Statistical Laboratory, Survey Section, Iowa State University, Ames, IA.

Horvitz, D.G., et Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association,* 47, 663-685.

Huang, E.T., et Fuller, W.A. (1978). Nonnegative regression estimation for sample survey data. *Proceedings of the Social Statistics Section*, American Statistical Association, 300-305.

Jagers, P. (1986). Post-stratification against bias in sampling. *International Statistical Review*/*Revue Internationale de Statistique,* 54, 159-167.

Kalton, G. (1983). *Compensating for Missing Survey Data.* Survey Research Center, University of Michigan, Ann Arbor, Michigan.

Kalton, G., et Kish, L. (1984). Some efficient random imputation methods. *Communications in Statistics,* A13, 1919-1939.

Kiaer, A. (1897). The representative method of statistical surveys (1976 English translation of the original Norwegian), Oslo, Central Bureau of Statistics of Norway.

Kim, J.K., et Fuller, W.A. (2004). Fractional hot deck imputation. *Biometrika,* 91, 559-578.

Kim, J.K., et Shao, J. (2013). *Statistical Methods for Handling Incomplete Data*. CRC Press, Boca Raton, FL.

Kish, L. (1995). The hundred years' wars of survey sampling. *Statistics in Transition*, 2, 813-830.

Kish, L., et Frankel, M.R. (1970). Balanced repeated replications for standard errors. *Journal of the American Statistical Association*, 65, 1071-1094.

Kish, L., et Frankel, M.R. (1974). Inference from complex samples (with discussion). *Journal of the Royal Statistical Society, Series B,* 36, 1-37.

Korn, E.L., et Graubard, B.I. (1995). Analysis of large health surveys: Accounting for the sampling designs. *Journal of the Royal Statistical Society, Series A,* 158, 263-295.

Kreuter, F. (2013). *Improving Surveys with Paradata.* Hoboken: Wiley.

Krewski, D., et Rao, J.N.K. (1981). Inference from stratified samples: Properties of the linearization, jackknife, and balanced repeated replication methods. *Annals of Statistics*, 9, 1010-1019.

Little, R.J.A. (1982). Models for nonresponse in sample surveys. *Journal of the American Statistical Association*, 77, 237-250.

Little, R.J.A., et Rubin, D.B. (1987, 2014). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons, Inc., (Deuxième édition 2014).

Madow, W.G. (1948). On the limiting distribution of estimates based on samples from finite universes. *Annals of Mathematical Statistics*, 19, 535-545.

Madow, W.G., et Madow, L.H. (1944). On the theory of systematic sampling, I. *The Annals of Mathematical Statistics*, 15, 1-24.

Madow, W.G., Nisselson, H. et Olkin, I. (Éds.) (1983). *Incomplete Data in Sample Surveys*, 1, 2, 3, New York: Academic Press.

Mahalanobis, P.C. (1939). A sample survey of the acreage under jute in Bengal. *Sankhyā*, 4, 511-531.

Mahalanobis, P.C. (1944). On large-scale sample surveys. *Philosophical Transactions of the Royal Society of London, Series B*, 231, 329-451.

Mahalanobis, P.C. (1946). Recent experiments in statistical sampling in the Indian Statistical Institute. *Journal of the Royal Statistical Society*, 109, 325-378.

McCarthy, P.J. (1966). Replication: An approach to the analysis of data from complex surveys. *Vital and Health Statisitcs*, Series 2, No. 14, National Center for Health Statistics, Public Health Service, Washington DC.

McCarthy, P.J. (1969). Pseudoreplication: Further evaluation and application of the balanced half-sample technique. *Vital and Health Statistics*, Series 2, No. 31, National Center for Health Statistics, Public Health Service, Washington, DC.

McCarthy, P.J. (1969). Pseudoreplication: Half-samples. *International Statistical Review*/*Revue Internationale de Statistique*, 37, 239-264.

McCarthy, P.J., et Snowden, L.B. (1985). The bootstrap and finite population sampling. *Vital Health Statistics*, 2-95, *Public Health Service Publication*, 85-1369, U.S. Government Printing Office, Washington, D.C.

Narain, R.D. (1951). On sampling without replacement with varying probabilities. *Journal of the Indian Society of Agricultural Statistics*, 3, 169-174.

Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97, 558-625.

Owen, A.B. (1988). Empirical likelihood ration confidence intervals for a single functional. *Biometrika*, 75, 237-249.

Pfeffermann, D. (2015). Methodological issues and challenges in the production of official statistics. *Journal of Survey Statistics and Methodology*, 3, 425-483.

Pfeffermann, D., et Sverchkov, M. (1999). Parametric and semiparametric estimation of regression models fitted to survey data. *Sankhyā, Series B,* 61, 166-186.

Porter, A.T., Holan, S.H., Wikle, C.K. et Cressie, N. (2014). Spatial Fay-Herriot model for small area estimation with functional covariates. *Spatial Statistics,* 10, 27-42.

Purcell, N., et Kish, L. (1979). Estimation for small domains. *Biometrics*, 35, 365-384.

Rao, J.N.K. (2003). *Small Area Estimation.* Hoboken: Wiley.

Rao, J.N.K. (2005). Évaluation de l'interaction entre la théorie et la pratique des enquêtes par sondage. *Techniques d'enquête,* 31, 2, 127-151. Article accessible à l'adresse http://www.statcan.gc.ca/pub/12-001-x/2005002/article/9040-fra.pdf.

Rao, J.N.K., et Molina, I. (2015). *Small Area Estimation: Second Edition.* Hoboken: Wiley.

Rao, J.N.K., et Scott, A.J. (1981). The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit and independence in two-way tables. *Journal of the American Statistical Association,* 76, 221-230.

Rao, J.N.K., et Scott, A.J. (1984). On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data. *Annals of Statistics,* 12, 46-60.

Rao, J.N.K., et Shao, J. (1999). Modified balanced repeated replication for complex survey data. *Biometrika,* 86, 403-415.

Rao, J.N.K., et Wu, C.F.J. (1988). Resampling inference with complex survey data. *Journal of the American Statistical Association,* 83, 231-241.

Royall, R.M. (1968). An old approach to finite population sampling. *Journal of the American Statistical Association,* 63, 1269-1279.

Royall, R.M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika,* 57, 377-387.

Rubin, D.B. (1974). Characterizing the estimation of parameters in incomplete data problems. *Journal of the American Statistical Association,* 69, 467-474.

Rubin, D.B. (1976). Inference and missing data. *Biometrika,* 63, 581-590.

Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys.* New York: John Wiley & Sons, Inc.

Schaible, W.L. (Éd.) (1996). *Indirect Estimators in U.S. Federal Programs.* New York: Springer.

Sitter, R.R. (1992). A resampling procedure for complex survey data. *Journal of the American Statistical Association,* 87, 755-765.

Skinner, C.J. (1994). Sample models and weights. Dans *Proceedings of the Survey Research Methods Section,* American Statistical Association, 133-142.

Steinberg, J. (Éd.) (1979). *Synthetic Estimates for Small Areas: Statistical Workshop Papers and Discussion.* NIDA Research Monograph nº 24. U.S. Government Printing Office, Washington, D.C., États-Unis.

Stephan, F., et McCarthy, P.J. (1958). *Sampling Opinions*. New York: John Wiley & Sons, Inc.

Sudman, S. (1966). Probability sampling with quotas. *Journal of the American Statistical Association,* 61, 749-791.

Sudman, S. (1976). *Applied Sampling.* New York: Academic Press.

Sukhatme, P.V. (1954). *Sampling Theory of Surveys with Applications*. Iowa State College Press, Ames.

Tam, S.-M., et Clarke, F. (2015). Big data, official statistics and some initiatives by the Australian Bureau of Statistics. *International Statistical Review*/*Revue Internationale de Statistique*, 83, 436-448.

Tchuprow, A.A. (1923). On the mathematical expectation of the moments of frequency distributions in the case of correlated observations. *Metron*, 2, 461-493, 646-683.

Thomsen, I. (1973). A note on the efficiency of weighting subclass means to reduce the effects of non-response when analyzing survey data. *Statistisk Tidskrift*, 11, 278-283.

van Remoortel, H., Giavedoni, S., Raste, Y., Burtin, C., Louvaris, Z., Gimeno-Santos, E., Langer, D., Glendenning, A., Hopkinson, N.S., Vogiatzis, I., Peterson, B.T., Wilson, F., Mann, B., Rabinovich, R., Puhan, M.A., Troosters, T. et PROactive consortium (2012). Validity of activity monitors in health and chronic disease: A systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, 9.

Wolter, K.M. (2007). *Introduction to Variance Estimation.* New York: Springer-Verlag.

Woodruff, R.S. (1952). Confidence intervals for medians and other position measures. *Journal of the American Statistical Association,* 47, 635-636.

Yates, F. (1948). Systematic sampling. *Philosophocal Transaction of the Royal Society of London, Series A,* A241, 345-377.

Yates, F. (1949). *Sampling Methods for Censuses and Surveys*. Griffin, Londres.