

Sur l'avenir de la statistique publique  
- perspectives pour la méthodologie

Carl-Erik Särndal  
Statistique Suède  
Université d'Örebro

Journées de méthodologie statistique  
INSEE, Paris  
24-26 janvier 2012

2012-01-20

## Le recensement du Canada 2011 - une cause célèbre

En 2010, le Gouvernement du Canada impose un nouveau régime pour la collecte de données pour le Recensement Canada 2011 (quinquennal).

Remplaçant une collecte systématique « un sur cinq », le gouvernement impose à Statistique Canada une *Enquête nationale auprès des ménages*, avec participation **volontaire** (« le formulaire long »).

# Le recensement du Canada 2011

Ceci malgré les protestations de la part de la communauté statistique au Canada.

Conséquence : débat animé dans les media.

L'opinion du Gouvernement : Les Canadiens ne devront plus être obligés de livrer des informations de nature parfois privée, faute de quoi risquer une amende.

Conséquences prévisibles:

Le plus nantis répondront à plus forte mesure ;  
les plus aisés, les mieux éduqués  
répondront plus que  
les groupes moins avantagés.

Sélection biaisée.

Vont alors faire défaut :

Des informations fiables,  
surtout sur ces derniers groupes.

# Le recensement du Canada 2011

Parmi les arguments: Il n'a pas été ***scientifiquement démontré*** que le nouveau système de collecte produira des chiffres moins précis qu'avant (le « un sur cinq systématique »)

Nous sommes donc invités à nous interroger sur  
« les bases scientifiques »  
de la statistique publique

Est-ce que la production de la statistique publique  
est tellement élémentaire,  
tellement dépourvue de théorie scientifique,  
qu'un peu de sens commun  
(ou même pas cela) suffira ?

Je m'interroge ...

D'ailleurs,  
a-t-on vraiment besoin de « chiffres précis » ?

Surtout quand, de nos jours,  
les sources d'erreur se multiplient,  
dans une production statistique  
de plus en plus complexe ?

Le rôle de la méthodologie ?

## L'inspiration de ma présentation

### Œuvres récentes :

Groves, R. M. and Lyberg, L. (2010) ***Total survey error, past, present, future***. Public Opinion Quarterly 24, 849-879

Kuusela, V. (2011) ***Paradigms in Statistical Inference for Finite Populations***. Thèse de doctorat (en statistiques), Université d'Helsinki.

Les deux traitent – de manières très différentes - de notions importantes pour la statistique publique

## Bibliographie, perspective historique

Desrosières, A. (1993) ***La politique des grands nombres ; histoire de la raison statistique.***  
Paris: La Découverte.

Bibliographie :

perspective théorie des sciences

Lakatos, I. (1970) ***Criticism and the Growth of Knowledge***. Cambridge Univ. Press.

Laudan, L. (1977) ***Progress and its problems. Towards a theory of scientific growth***. Univ. of California Press.

Bohm, D. and Peat, F.D. (2000) ***Science, Order and Creativity***, 2<sup>nd</sup> edition. London: Routledge.

Bibliographie :

les doutes , les signes inquiétants

Groves, R. (1987) ***Survey research is a methodology without a unifying theory.*** Public Opinion Quarterly, 51, 156-172.

Franchet, Y. and Nanopoulos, P. (1997) ***Statistical science and the European statistical system.*** In Proc. Conference in honour of S. Franscini. Basel: Birkhäuser.

Platek, R. and Särndal, C.E. (2001) ***Can a statistician deliver ?*** Journal of Official Statistics 17, 1-127, with 16 discussions

## Structure de mon intervention

1. La **statistique publique** : Possède-t-elle une théorie (scientifique) ? La réponse : Non .
2. La **science statistique** vis-à-vis  
La production de la statistique publique :  
Quelle relation, quel impact ?
3. La **méthodologie** de la statistique publique est un **champ fragmenté**.
4. La **théorie des enquêtes** (par sondage)  
vis-à-vis La **méthodologie de la statistique publique**  
Quelle relation ?
5. Le **raisonnement probabiliste** et son **insuffisance**

# 1. La statistique publique : A-t-elle une théorie ?

Groves (1987) : ***Survey research is a methodology without a unifying theory.***

un titre révélateur ...

S'il y avait une théorie solide, alors des incidents regrettables comme « Recensement Canada 2011 » ne se produiraient pas.

Des percées théoriques impressionnantes et prestigieuses sont moins visibles aujourd'hui.

Comparer l'acceptation, les années 1930, de l'échantillonnage (plutôt que le dénombrement total)

## « Collection de pratiques »

Statistique Canada (1998) : Les techniques d'enquête (le « survey methodology ») sont une collection de pratiques étayées par des notions théoriques qui s'offrent aux praticiens ... il appartient à ces derniers de les choisir avec soin, en vue de répondre à leurs besoins particuliers.

Il n'y a pas de code précis.

Ainsi, de l'extérieur, on a tendance à regarder la production de la statistique publique comme un « boulot utilisant des techniques assorties »

alors, disent-ils, peu importe les techniques utilisées, dans le recensement, dans les enquêtes (au Canada ou ailleurs); on aura ***des chiffres***.

## Un rêve

Groves (1987) formule ainsi le rêve (ou la nostalgie) d'une théorie achevée :

« Une théorie des enquêtes rejoindrait des concepts des sciences du comportement avec les propriétés statistiques des estimations »

La ***science statistique***, elle, par contre, possède une théorie fort bien structurée

## 2. La science statistique vis-à-vis la production de la statistique publique

Idée centrale de la *science statistique* :

A partir d'une partie (un *échantillon*),  
se prononcer sur la totalité (la *population*)

C'est *l'inférence statistique*, assistée de méthodes  
(intervalles de confiance, tests d'hypothèses),  
et ceci à un *niveau quantifiable de probabilité*.

Mot clef : la ***probabilité***, concept de base pour  
*l'inférence statistique* ,  
laquelle est une *théorie mathématique*

L'inférence statistique a connu un développement impressionnant les années 1930 et 40 ;  
les racines sont cependant à chercher plus loin  
(exemple : P.S. de Laplace vers 1810)

Pour la théorie des enquêtes par sondage :  
J. Neyman (années 1930).

Voir : thèse de Kuusela (2011) ; Desrosières (1993)

*La statistique publique* nous donne

« des chiffres sur la population »

ceci contrastant avec ce que préconise  
la *science statistique*, à savoir

« des inférences sur la population »

La statistique publique se trouve dans l'impossibilité  
de fournir des énoncés sur la population  
à un niveau quantifiable de probabilité

La science statistique vis-à-vis  
La production des statistiques officielles

A ce point de ma présentation, certains se sentent mal à l'aise : Que dites-vous là ? Pas d'inférence sur la population ? Bien sûr on fait des inférences !

Ma réponse : Mais non ! En statistique publique, on ne se prononce pas, probabilités à l'appui, sur la population.

Par contre, des chiffres il y en a, en grande quantité.

Les responsables s'évertuent de dire, peut-être :  
« nos chiffres possèdent la meilleure qualité possible »

Mais « qualité », ça, c'est autre chose.

La science statistique vis-à-vis  
La production de la statistique publique

Franchet et Nanopoulos (1997) : « La méthodologie de la statistique publique est une notion qui s'interprète à partir de la notion de méthodologie de *la statistique mathématique* .... *Le formalisme probabiliste* de la statistique mathématique a fourni à la statistique publique le cadre nécessaire pour son fondement scientifique »

Belle déclaration, plus ou moins réaliste.

Mon point de vue: un contraste frappant  
entre d'un côté :

***les principes de la science statistique*** -  
mathématiquement fondés

de l'autre côté :

***la production de la statistique publique***  
une collection, un nuage de pratiques

Si

***la production de la statistique publique***

ne possède pas une théorie solide et unifiée,

elle possède cependant ce qu'on appelle

***une méthodologie***

Mais c'est un champ d'étude ***fragmenté***

### 3. La méthodologie de la statistique publique : un champ fragmenté

***Fragmentation*** d'une science, d'un champ d'étude :

Quand des divisions se produisent dans une manière plus ou moins arbitraire sans égard pour un contexte qui est, par sa nature même, plus vaste

Référence:

D. Bohm et F.D. Peat (2000) Science, order and creativity

## La fragmentation

Deux aspects:

(i) Des théories concurrentes créent des divisions

(ii) « Théorie par de petits bouts » s'établit  
dans un champ d'étude ayant pourtant  
besoin de cohérence

Signe de fragmentation d'une science :

La formation de groupes de chercheurs, unis par un intérêt commun d'étudier un problème (d'envergure plus ou moins restreinte).

Des groupes de personnes se mettent ensemble, souvent sous un nouveau nom, une rubrique qui n'existait pas il y a 10 ans

C'est naturel, peut-être: Dans un champ d'étude, de nouveaux problèmes surgissent ; exigent une attention particulière ; un bout de théorie se développe autour du problème.

Résultat: des spécialisations en dedans d'un champ d'étude qui est par sa nature même plus vaste , ayant besoin de cohérence

Ainsi la statistique publique est arrivée à reposer sur « une collection de pratiques étayées par des notions théoriques » par ici par là

La statistique publique comporte bon nombre d'exemples de fragmentation, des « sous-disciplines ».

Relativement au traitement des données:

Imputation

Traitement de la non-réponse

Vérification (editing)

Estimation pour les petits domaines

Relativement au collecte des données:

Méthodes concernant

Fardeau des répondants

Motiver les répondants

Confidentialité

et d'autres

Ainsi, dans la méthodologie de la statistique publique,  
des groupes actifs de recherche,  
des « networks », des spécialisations

Bon pour la discipline ?

Pour le long terme, quelles conséquences ?

« Une vision à long terme, encadrant l'ensemble des idées et notions, est d'importance primordiale lors du développement d'un champ d'étude ; on n'y arrive pas par des spécialisations étroites. »

Bohm et Peat (2000)

La méthodologie de la statistique publique ne possède pas une vision à long terme.

#### 4. La *théorie des enquêtes* (par sondage) vis-à-vis la *méthodologie de la statistique* *publique*

Dans une perspective historique,  
une distinction s'impose :

*D'une part, la méthodologie* – l'ensemble d'outils  
pratiqués aujourd'hui dans la statistique publique

*D'autre part, la théorie des enquêtes par sondage*  
– champ mathématique, basé sur un concept  
central de la science statistique:

A partir *d'une part*, arriver aux conclusions  
*probabilistes* sur *le tout*.

## La théorie des sondages

- une théorie mathématique (d'inférence)
- qui, dans ses meilleurs manifestations, sur cent ans et plus, a grandement influencé la pratique
- est malgré cela enseignée dans peu d'universités

Témoin: Livre « IASS Landmark papers in Survey Statistics » : 19 articles, presque tous de caractère mathématique

Thèse de Kuusela: Long chemin :P.S. Laplace (1810)  
à Cochran, Deming et Hansen (années 1950)

## La théorie des sondages

Les classiques :

W.G. Cochran, W.E. Deming, M.H. Hansen ,

des mathématiciens

avec une profonde compréhension

des exigences pratiques

## La théorie

Deming (1960) :

« Un statisticien théorique est quelqu'un qui se laisse guider, dans sa pratique, par la théorie.

Le statisticien théorique, c'est l'homme pratique, car il possède un meilleur guide pour la pratique que les erreurs commises par ses prédécesseurs.

La théorie statistique montre comment les mathématiques, le bon jugement, et les connaissances du sujet travaillent ensemble »

Une belle déclaration ... j'en conviens

## La théorie des sondages

un acheminement, sur plus de cent ans, en bonne partie depuis 1950. Est-elle arrivée à **maturité** ?

Imre Lakatos (théoricien des sciences) :

« Une science est arrivée à la maturité quand les scientifiques dans ce champ d'étude ignorent carrément les anomalies et les influences intellectuelles et sociales venant de l'extérieur, pour concentrer entièrement leurs efforts sur les articulations mathématiques du programme de recherche »

## Science arrivée à maturité

Lakatos: Les chercheurs « ignorent carrément les anormalités et les influences ... venant de l'extérieur, pour concentrer entièrement leurs efforts sur les articulations mathématiques »

Une goutte d'ironie : On arrive à faire publier ses travaux dans des revues scientifiques , même si le contenu n'a rien à voir avec la réalité pratique.

Le problème de « la mathématisation »

La théorie des sondages est arrivée à maturité

Aspect positif : Une présence incontournable sur la scène des sciences

Ses meilleures manifestations ont grandement influencé la pratique

Et on arrive à faire publier ses travaux théoriques, même ceux qui n'ont rien à faire avec la pratique.

Par contre, la réalité aujourd'hui fait appel à  
La méthodologie (de la production statistique)

Cochran , Deming et Hansen : Si l'on feuillette  
leurs livres classiques des années 1950,

Survey methodology : Le terme n'est pas là.

Imputation, small area estimation, editing : Absents  
eux aussi

Non-réponse : mentionné en passant

Groves et Lyberg (2010) notent :

Le livre de Hansen, Hurwitz et Madow (1953),  
638 pages, dont 9 pages consacrées à « réponse  
et autres erreurs non-dues à l'échantillonnage »

Les livres classiques (Cochran, Deming, Hansen)  
visent entièrement les propriétés de l'erreur due à  
l'échantillonnage (par la théorie alors en plein  
essor).

## La méthodologie

Groves et Lyberg expliquent ainsi cette *vision étroite* concentrée sur l'erreur due à l'échantillonnage :

Laisser pour compte les autres erreurs n'était pas surprenant à l'époque (années 1950) car

l'échantillonnage (plutôt que le dénombrement total) était nouveau, pas encore accepté;

il fallait en faire la promotion,

le mettre de l'avant,

le défendre par tous les moyens

Mon interprétation est différente : Si on avait demandé à Cochran, à Deming ou à Hansen, vers 1955 :

« Que pensez vous de la méthodologie aujourd'hui (c'est-à-dire en 1955) »

Ils n'auraient pas su bien répondre :

Méthodologie des enquêtes par sondage, c'est quoi? Survey methodology , c'est quoi ?

Le terme n'était pas dans le langage courant les années 1950.

Survey (sampling) theory : Oui, ça par contre existait en 1955

« La méthodologie » est une notion moderne,

engendrée par

la complexité toujours grandissante  
des enquêtes statistiques modernes.

## La méthodologie

Ainsi nous sommes en présence d'une « collection de pratiques étayées par des notions théoriques » dans laquelle il faut « choisir avec soin » en vue de répondre à ses besoins particuliers.

Mais il n'y a pas de code précis.

Ce nuage flou de « pratiques et techniques » s'est érigé au delà de la ***théorie des enquêtes (par sondage)***, champ mathématique, d'une ancienneté de > 100 ans.

**La *méthodologie des enquêtes statistiques***  
est une discipline dont les ingrédients,  
aujourd'hui, proviennent de plusieurs sciences

non seulement la *Science statistique*

mais aussi

*Psychologie*

*Sociologie*

*Sciences économiques*

et surtout

*Informatique*

## L'apprentissage de la méthodologie :

Les programmes complets de formation  
sont rares, mais ils existent :

ENSAI en France

JPSM aux E.-U.

OM

## La méthodologie

Ce genre de programme, là où il existe, offre une grande variété de cours:

Modes de collecte de données, comportement des répondants, interviewing, pre-testing, fardeau des répondants, confidentialité, etc.

Cours disparates – quoique importants :  
Signe d'un champ fragmenté

OM

Ainsi, pour ***l'enseignement et la formation*** :

***Science statistique*** : dans toutes les universités

***Théorie des enquêtes par sondage*** : dans quelques rares universités

***Méthodologie des enquêtes*** : dans quelques centres spécialisés, mais pratiquée partout, plus ou moins bien (dans toutes les agences nationales)

## 5. Raisonnement probabiliste et son insuffisance

La *production de la statistique publique* est « passée outre » la *science statistique*.

La ***probabilité***, concept fondamentale de la science statistique, est une notion « trop étroite » pour la statistique publique.

La statistique publique, par sa complexité, refuse de s'intégrer dans cette enveloppe trop étroite.

*La probabilité et l'erreur probable*, éléments pourtant fondamentaux de la science statistique, jouent un rôle minime quand le grand public regarde les chiffres publiés.

Pour eux, les chiffres, c'est la vérité.

Franchet et Nanopoulos (1997) : « Très souvent, ou presque toujours, les résultats sont présentés comme étant la vérité absolue, exprimée par des chiffres ayant l'air exacts ... Des intervalles de confiance ne sont pas fournis, les méthodes d'estimation ne sont pas présentes ... des tests de signification ne sont pas donnés »

## L'insuffisance du cadre probabiliste

J'ai l'impression que Franchet et Nanopoulos (1997) disent : Il faut que nous, Eurostat et les NSI, nous réussissions, dans un avenir proche, de pouvoir fournir les expressions de précision, les intervalles de confiance au niveau probabilité 95%

Le fardeau de la preuve repose sur nous, les statisticiens des NSI.

Belle illusion sur l'avenir ... même si on insiste sur la nécessité de l'exactitude

dans les statistiques qui forment la base des décisions cruciales pour la nation.

La responsabilité du statisticien ?

Can a statistician deliver ?

(Quelle est la responsabilité du statisticien de livrer des statistiques « justes » ?)

Article JOS 2001 de R. Platek et C.E. Särndal

Suivi de 16 discussions.

Les 16 discussions :

Certains disent : Il est évident que nous n'aurons jamais une « théorie parfaite », le processus derrière cela est beaucoup trop complexe.

D'autres disent: « the glass is more than half full »

Admettons: L'article *Can a statistician deliver* était axé sur l'optique « science statistique », déplorant l'impossibilité de circonscrire l'erreur totale d'un sondage par le concept de probabilité.

Mais dites-vous, nous avons la théorie (le paradigme)  
de *l'erreur totale des statistiques d'enquête*  
(le «total survey error») dont parlent Groves et Lyberg.

A notre grande déception : Ce paradigme (vieux d'au  
moins 30 ans) n'a pas réussi à assurer, de façon  
routinière et automatisée, des mesures (des  
indicateurs) des propriétés statistiques (biais,  
variance) provenant de *toutes les causes*

Néanmoins, comprenez moi bien,  
nous, les méthodologistes, formons  
une catégorie très importante

et le raisonnement statistique joue un rôle clef  
dans la production de la statistique :

modélisations de toute sorte, les lois statistiques,  
les séries chronologiques, la confidentialité,  
l'administration du fardeau des répondants ...

Et pourtant, si je retourne à mon point de départ,  
comment est-ce possible que cela arrive,  
l'incident regrettable du Recensement Canada 2011  
et d'autres interventions non-sollicitées du même  
genre ?

## L'avenir de la méthodologie

« The show must go on »

car il y aura toujours

- des nouveaux besoins d'information
- sur des nouveaux groupes dans la société
- avec des nouvelles possibilités de collecte de données.

«Les besoins», formulés par d'autres («la société»),  
dominent le travail des méthodologistes.

OM

## L'avenir de la méthodologie

Les méthodologistes :

toujours soucieux de mesurer avec ***exactitude***  
des concepts et des notions

dont l'importance a été prescrite par d'autres;

une ***exactitude*** qui leur échappe toujours,  
à leur grande déception.

## References

- Bohm, D. and Peat, F.D. (2000). *Science, Order, and Creativity*, 2<sup>nd</sup> edn. London: Routledge.
- Desrosières, A. (1993) *La politique des grands nombres ; histoire de la raison statistique*. Paris: La Découverte.
- Franchet, Y. and Nanopoulos, P. (1997). *Statistical science and the European statistical system: Expectations and perspectives*. In Proc. Conference in honour of S. Franscini. Basel: Birkhäuser.
- Groves, R. (1987). Survey research is a methodology without a unifying theory. *publique Opinion Quarterly*, 51, 156-172.
- Groves, R. M. and Lyberg, L. (2010) Total survey error, past, present, future. *publique Opinion Quarterly* 24, 849-879
- Kuusela, V. (2011) *Paradigms in Statistical Inference for Finite Populations*. Thèse de doctorat (statistique), Université d'Helsinki.
- Lakatos, I. (1970). A chapter in: *Criticism and the Growth of Knowledge*. Cambridge Univ. Press.
- Laudan, L. (1977). *Progress and its Problems. Toward a theory of scientific growth*. LA: Univ. of California Press.
- Platek, R. and Särndal, C.E. (2001). Can a statistician deliver? *J. Official Statistics*, 17, 1 – 127 (with 16 discussions)