



N° 11-522-XIF au catalogue

**La série des symposiums internationaux
de Statistique Canada - Recueil**

Symposium 2003 : Défis reliés à la réalisation d'enquêtes pour la prochaine décennie

2003



Statistique
Canada

Statistics
Canada

Canada

Recueil du Symposium 2003 de Statistique Canada
Défis reliés à la réalisation d'enquêtes pour la prochaine décennie

STATISTIQUE OFFICIELLE : LES PROCHAINS DÉFIS MÉTHODOLOGIQUES

D. Holt¹

RÉSUMÉ

La méthodologie joue un rôle important dans l'assurance de la qualité des produits statistiques individuels. Mais elle a aussi une fonction stratégique plus générale au sein de tout bureau national de la statistique (BNS), à savoir promouvoir la réputation professionnelle de l'organisme, contribuer à l'élaboration de normes internationales et promouvoir la confiance du public dans le BNS et ses produits. Donc, quand on examine les orientations futures de la recherche et du développement méthodologiques, il est essentiel de tenir compte des besoins stratégiques du BNS proprement dit. Le présent article vise à illustrer le lien entre les besoins stratégiques du BNS et les besoins méthodologiques qui en découlent.

MOTS CLÉS : Estimation basée sur un modèle, estimation des petites régions, priorités méthodologiques, stratégie globale, surveillance du rendement.

1. INTRODUCTION

Préciser les futurs défis que posera le développement méthodologique dans le domaine de la statistique officielle n'est pas un exercice facile. Dans de nombreux cas, les développements méthodologiques de court terme s'inscrivent naturellement dans la foulée des programmes de développement existants, mais il n'en n'est pas de même des projets de développement de long terme. Au moment où les nouveaux besoins statistiques sont énoncés correctement, il est souvent un peu trop tard pour pouvoir mettre au point en temps voulu la méthodologie de soutien. De surcroît, tout effort en vue de préciser les besoins méthodologiques futurs peut aboutir à l'établissement d'une liste assez peu structurée, dont ne se dégage aucune impression d'orientation globale. Le présent article ne consistera pas en une revue exhaustive des champs futurs de développement. Il visera plutôt à énoncer un processus destiné à relier les besoins stratégiques du bureau national de la statistique au programme de développement méthodologique.

Qu'est-ce que la méthodologie?

La mission de tout bureau national de la statistique (BNS) consiste à produire des statistiques qui répondent aux besoins des groupes d'utilisateurs, dont la qualité est de haut niveau et défendable (en ce qui concerne les principales utilisations des résultats statistiques) et qui inspirent confiance aux membres du public. Cette dernière exigence est fondamentale, car, si les statistiques n'inspirent pas confiance au public, il est difficile de les utiliser pour :

- élaborer des politiques et en faire le suivi;
- gérer les services publics;
- faciliter les prises de décision des fonctionnaires et des gens d'affaires;
- appuyer d'autres utilisations comme l'utilisation à des fins d'éducatives;
- fournir au public l'information nécessaire pour surveiller les activités du gouvernement proprement dit, en tant que pierre angulaire du processus démocratique.

Pour nombre d'entre nous, jusqu'à présent, les activités de développement méthodologique ont eu pour noyau la théorie de l'échantillonnage (y compris les propriétés des bases de sondage), l'estimation statistique, la

¹ D. Holt, University of Southampton. L'auteur était anciennement directeur de l'Office for National Statistics (ONS) et chef du Government Statistical Service du Royaume-Uni.

méthodologie d'enquête et les méthodes statistiques reliées à la saisie et au traitement des données (comme la vérification et l'imputation, et la détection des valeurs extrêmes). Plus récemment, la modélisation statistique a joué un rôle d'importance croissante dans la résolution de problèmes, comme l'estimation des petites régions et l'effet de l'erreur de mesure sur les estimations. Naturellement, l'utilisation de modèles statistiques (explicites ou implicites) en statistique officielle remonte beaucoup plus loin dans le temps et des problèmes tels que la correction pour la non-réponse, l'analyse de séries chronologiques et la désaisonnalisation, qui peuvent tous être considérés comme ayant une solution implicitement ou explicitement fondée sur un modèle, sont étudiés de façon très générale depuis longtemps.

En réalité, une définition nettement plus générale de la méthodologie s'impose. Le cadre complet des comptes nationaux, l'élaboration de comptes satellites pour des domaines particuliers (p. ex. les comptes de l'environnement, de la main-d'œuvre et de la santé) et des sujets tels que l'analyse des entrées-sorties pour appuyer le processus d'équilibrage des comptes nationaux sont des domaines où le développement méthodologique a une place.

De façon plus générale, le cadre de production des statistiques officielles repose sur l'énoncé des concepts que nous cherchons à mesurer et, à partir de ces concepts, sur l'élaboration de définitions, puis des mesures statistiques qui appuieront chaque produit statistique. La mise en place de ce cadre de travail comprend une composante méthodologique, même si nombre d'aspects nécessitent la participation d'autres employés du BNS et des utilisateurs externes. Donc, le cadre de définitions, de classifications et de meilleures pratiques qui, ensemble, constituent les normes internationales fait partie de ma définition générale de la méthodologie.

En tant que méthodologistes, nous nous efforçons d'établir des principes et des pratiques, des algorithmes et une théorie générale applicables à diverses situations et, par conséquent, de fournir un cadre de référence général dans lequel sont établis les systèmes que nécessite la production de statistiques officielles. L'élaboration du plan de sondage, la stratification a posteriori, les systèmes d'estimation par la régression généralisée (GREG), la vérification et l'imputation automatique et la désaisonnalisation sont tous des exemples de méthodes généralisées. Mais l'expérience professionnelle qui assure que ces approches généralisées soient utilisées de façon sensée dans le cadre de chaque application, puis que les résultats soient évalués comme il convient est un élément tout aussi important.

L'incidence de la méthodologie

Des bases méthodologiques solides sont indispensables à la concrétisation de tous les aspects de la mission fondamentale du BNS, c'est-à-dire la production de statistiques pertinentes, de bonne qualité, inspirant confiance aux membres du public. Le rôle de la méthodologie ne se limite pas à appuyer individuellement chaque produit statistique, mais est, au contraire, plus général.

Le rôle fondamental d'une bonne méthodologie en tant que pierre angulaire de chaque produit statistique est bien établi. À ce niveau, les utilisateurs y comptent bien : le besoin de méthodes justifiables, de grande qualité, et de marges d'erreur connues et publiées afin que les utilisateur puissent interpréter les résultats est bien compris et je ne m'étendrai pas davantage sur ce sujet ici.

Mais le rôle d'une bonne méthodologie va au-delà de son incidence sur un produit particulier et consiste à soutenir la mission du BNS tant en ce qui concerne la façon dont l'organisme fonctionne que, de façon plus générale, la confiance du public dans l'organisme et ses produits. Les effets d'une méthodologie solide sont triples :

- elle favorise l'implantation, à l'échelle du BNS, d'une culture d'évaluation et d'examen systématiques qui est essentielle au maintien de normes de qualité;
- elle contribue, notamment grâce à l'enracinement de ce genre de culture, à susciter la confiance des utilisateurs dans le professionnalisme du BNS et, par conséquent, dans ses produits individuels;
- elle soutient la recherche et le développement nécessaires à la mise au point de nouvelles statistiques ou à l'examen et au renouvellement des statistiques existantes afin d'assurer que les produits du BNS continuent de répondre aux besoins des utilisateurs dans l'avenir.

Étant donné le rôle crucial de la méthodologie dans la réalisation de la mission du BNS, on ne peut considérer la question des futurs défis méthodologiques que dans le contexte des futurs besoins stratégiques du BNS proprement dit. Aussi les priorités pourraient-elles varier d'un pays à l'autre, mais, étant donné les problèmes généraux auxquels

doivent faire face les BNS, il existera bien des points communs. Plus précisément, les orientations stratégiques que devrait suivre le développement méthodologique ne peuvent se dégager uniquement d'une évaluation interne réalisée par le groupe de la méthodologie proprement dit.

2. CONTEXTE STRATÉGIQUE

La présente section brosse un tableau succinct de certaines pressions stratégiques qui s'exercent sur le BNS. Ces pressions influent fortement sur les futurs défis que posera le développement méthodologique.

Maintenir la pertinence et la qualité des produits courants

Les défis les plus immédiats que doit relever le BNS consistent sans doute à maintenir la pertinence et la qualité des produits statistiques existants. Ces derniers doivent être examinés et révisés à mesure qu'évoluent les conditions socioéconomiques au sein de la société. Dans l'ensemble, cet exercice demande la mise à jour de la méthodologie sous-jacente, mais peut aussi nécessiter l'élaboration de nouvelles mesures statistiques pour refléter les nouveaux aspects de la société. Les domaines dans lesquels des changements ont eu lieu et se poursuivent sont bien connus. En voici quelques exemples.

- Mondialisation de l'économie : l'interdépendance économique croissante des pays, la mise en place et l'exécution d'activités économiques transfrontalières constituent une part de plus en plus importante de l'activité économique mondiale. Il n'est pas inhabituel que des entreprises établissent leur siège social dans un pays, installent des capacités de conception dans un autre, réunissent des capitaux d'investissement dans un troisième et établissent des unités de fabrication ou de service dans d'autres. Les facteurs fondamentaux de la production, y compris la main-d'œuvre, peuvent être répartis à travers le monde. Ce fait n'est pas nouveau, mais la croissance régulière de ce genre d'activités complique encore davantage la tâche des statisticiens officiels qui doivent mesurer de façon fiable les activités dans les comptes nationaux.
- La mondialisation est stimulée en grande partie par les nouvelles technologies de l'information et des communications (TIC) et la façon dont elles permettent de restructurer les activités commerciales afin de profiter des avantages économiques. Les éléments sous-jacents sont la croissance du secteur des TIC et, de façon plus générale, la pénétration importante de ces technologies dans presque tous les aspects de l'économie et de la société. L'accès à l'information et l'usage actif de cette dernière ont des retombées sur presque tous les aspects de l'industrie et du commerce, des services publics, comme la santé et l'éducation, et plus généralement, du gouvernement. Ils modifient aussi la façon dont les citoyens vivent leur vie et dont ils ont accès à l'information et l'utilisent à des fins de consommation, d'interaction avec l'économie et les services publics, de perfectionnement et de loisirs. Donc, certains changements ont une incidence sur les aspects offre et demande de l'économie et d'autres dépassent le cadre de la production et touchent toutes les facettes de la société. Un très grand nombre de pays ont adopté des politiques qui favorisent le développement des TIC et l'accès du public à ces dernières (ladite société de l'information) afin d'exploiter les possibilités de croissance économique et de concurrence. Les statisticiens officiels doivent suivre l'évolution de la société et surveiller l'efficacité des politiques publiques conçues pour appuyer cette évolution.

Dimension géographique

La demande de statistiques répondant aux exigences des utilisateurs aux niveaux supra et infranational est de plus en plus pressante. Dans le premier cas, la pression résulte principalement du besoin de statistiques comparables sur le plan international, à des fins aussi bien nationales qu'internationales. Ces statistiques sont utilisées par les gouvernements nationaux afin de surveiller la performance de leur pays comparativement à leurs concurrents, pour s'assurer du maintien ou de l'amélioration de leur compétitivité économique ou pour surveiller les progrès économiques et sociaux dans d'autres pays, ainsi que le résultat d'autres politiques économiques et sociales que pourraient adopter d'autres nations.

De plus en plus, certains pays utilisent ces statistiques en vue de participer à des processus internationaux de prise de décision et de répartition des ressources. Dans de telles circonstances, il est essentiel de disposer de statistiques comparables à l'échelon international.

Les organismes internationaux ont également besoin de ce genre de statistiques pour surveiller les rendements nationaux et faire des comparaisons. La Banque mondiale, le FMI et les organismes de financement bilatéraux dépendent fortement des statistiques officielles pour suivre l'effet des politiques et des programmes d'aide technique. Par exemple, un examen des sommets et des grandes conférences tenus par les Nations Unies durant les années 1990 a permis de cerner plus de 280 indicateurs statistiques jugés nécessaires pour surveiller l'application des politiques faisant suite aux décisions prises lors de ces conférences (CSONU, 2002).

On pourrait penser que la statistique économique est le domaine où le besoin de données comparables à l'échelon international est le plus prononcé, mais il existe un intérêt marqué au niveau régional ainsi qu'international dans les domaines de l'environnement, de la santé, de l'éducation et de la culture.

Parallèlement, la demande de statistiques infranationales en vue d'élaborer diverses politiques ne cesse de croître. Plusieurs pays ont de plus en plus tendance à décentraliser les pouvoirs et à transférer certaines responsabilités en matière d'élaboration de politique et de prise de décision à une forme ou l'autre de gouvernement régional. À un palier inférieur, il existe une demande croissante de statistiques visant des régions beaucoup plus petites (les dites statistiques de quartier) en vue de surveiller l'effet des politiques sociales et économiques sur les petites collectivités. Cette demande est souvent liée à des préoccupations concernant le désavantage social et les privations multiples.

À ces faits nouveaux, il faut ajouter un phénomène apparenté, c'est-à-dire un intérêt croissant pour des solutions éventuelles de remplacement du recensement traditionnel, grâce à l'utilisation de registres administratifs, que les pays nordiques ont été les premiers à expérimenter, ou à l'utilisation d'enquêtes continues à grande échelle (approche dite du recensement continu).

Sources administratives

La demande de statistiques infranationales pousse de plus en plus à utiliser des sources de données administratives (peut-être conjuguées à des données d'enquête) en vue d'améliorer les estimations. Elle met aussi de nouveau l'accent sur les travaux réalisés dans le domaine de l'estimation des petites régions. La désagrégation plus poussée des données pose des problèmes de confidentialité et de divulgation. Les pressions visant à une utilisation plus complète des sources administratives tiennent aussi à la création de bases de données détaillées plus facilement consultables grâce aux TIC et à la possibilité de réduire le fardeau de réponse grâce à l'utilisation de l'information existante. À mesure que les TIC transformeront les systèmes administratifs utilisés par les gouvernements pour gérer les services publics de grande envergure, comme les services fiscaux, de protection sociale, de santé et d'éducation, il faudra s'attendre à ce que les possibilités et l'obligation d'utiliser ces sources de données à des fins statistiques augmentent.

Ni la pression ascendante en vue de produire des statistiques comparables à l'échelon international, ni la pression descendante en vue de produire des statistiques infranationales n'est un fait nouveau, mais l'une et l'autre se maintiennent et indiquent que les BNS ont encore beaucoup à faire pour répondre aux besoins des utilisateurs.

Mesure du rendement

Dans plusieurs pays, la demande de statistiques permettant de surveiller le rendement des services publics au niveau d'unités de prestation identifiables s'accroît. Cette croissance émane partiellement de la volonté d'obliger davantage les administrations publiques à rendre compte du rendement des services publics et partiellement de l'importance grandissante accordée aux utilisateurs des services publics en tant que consommateurs ayant besoin d'information leur permettant d'avoir efficacement accès à ces services. Donc, il existe une tendance croissante à surveiller et à publier des paramètres tels que le taux de rendement scolaire dans les écoles, les mesures du rendement des hôpitaux, le taux d'affaires criminelles tirées au clair par les services de police, voire même le taux de survie des patients de chirurgiens cardiaques individuels.

Toutes ces tendances et ces faits nouveaux ont pour fondement l'obligation qu'ont les BNS de peindre un tableau pertinent, cohérent et digne de foi de la situation socioéconomique et de justifier ainsi la confiance permanente des membres du public dans les produits statistiques.

À la fin de l'article, je m'efforce de préciser quelles seront les implications des besoins stratégiques susmentionnés en ce qui concerne les futurs développements méthodologiques. L'énoncé de ces implications sera inévitablement partiel et d'autres besoins en matière de développement pourraient être identifiés.

3. CADRE THÉORIQUE DE L'ÉCHANTILLONNAGE

La théorie fondamentale de l'échantillonnage bien connue sur laquelle repose la plupart de la statistique officielle s'inspire des travaux de Neyman (1934). Supposons que nous ayons une population finie fixe de taille N comprenant les valeurs unitaires $Y_i, i = 1, \dots, N$. À partir de cette population, nous sélectionnons un échantillon, s , comprenant les valeurs $Y_i, i \in s$ selon un plan d'échantillonnage aléatoire p_s . L'inférence est fondée sur la loi aléatoire générée par le plan d'échantillonnage.

Le grand avantage de cette approche est que les statisticiens officiels peuvent défendre la thèse de l'absence de biais dans les estimations, parce que celles-ci ne reposent sur aucune hypothèse. Si nous supposons que la base de sondage et la réponse sont complètes, alors, quel que soit, par exemple, le plan de sondage stratifié à plusieurs degrés (y compris le choix de la stratification et des variables auxiliaires, des limites de strates et de la répartition de l'échantillon), la loi aléatoire assurera que l'estimateur convergent par rapport au plan de sondage soit sans biais (ou le soit effectivement si nous nous intéressons à des estimateurs par le quotient ou par la régression et à des échantillons raisonnablement grands). Si le plan de sondage est bon et que le choix des variables de stratification, de la taille des grappes, des variables auxiliaires et de la répartition de l'échantillon est raisonnable, les estimations seront essentiellement sans biais par rapport au plan d'échantillonnage et d'une efficacité raisonnable. Si les choix sont médiocres, les estimations seront encore sans biais par rapport au plan d'échantillonnage, mais leur efficacité sera moins bonne.

Il s'agit là d'un aspect qui justifie la confiance du public et cela a du bon. Il convient de souligner que la défense fondamentale des produits statistiques n'est pas que les statisticiens officiels sont hautement compétents et professionnels, mais qu'ils utilisent des méthodes qui excluent spécifiquement les jugements professionnels susceptibles d'introduire un biais dans les estimations résultantes. Dans la mesure du possible, nous nous sommes accrochés résolument à ce principe : par exemple, nous avons créé un cadre général de production d'estimations essentiellement convergentes par rapport au plan de sondage, y compris les estimations GREG. Cette approche a bien servi les statisticiens officiels, a produit des statistiques défendables et, donc, suscité la confiance des membres du public dans les BNS et leurs travaux.

Naturellement, à la limite, cela a toujours été un peu une illusion, puisque les méthodes de correction pour la non-réponse, l'imputation pour remplacer les valeurs manquantes et les procédures particulières, comme la désaisonnalisation des séries chronologiques, sont toujours justifiées par un ensemble d'hypothèses sous-jacentes explicites ou implicites.

Cependant, il est évident depuis un certain temps que, dans des circonstances particulières (p. ex. l'estimation des petites régions) où le prix de l'absence de biais est trop élevé, l'utilisation d'estimateurs fondés sur un modèle ne cesse de croître.

Je soutiendrai, en temps voulu, qu'il existe des situations où les estimations fondées sur un modèle ne sont pas simplement inévitables, mais font partie intégrante de l'objectif et que cet état de chose soulève plusieurs questions explicites auxquelles doivent répondre les méthodologistes des BNS. Mais, pour commencer, je m'attarderai à un ou deux aspects de l'estimation sur petits domaines pour lesquels le niveau de développement méthodologique semble encore ne pas répondre aux besoins stratégiques des BNS.

4. ESTIMATION DES PETITES RÉGIONS

Les travaux méthodologiques portant sur ce sujet, qui remontent assez loin dans le temps, ont donné lieu à une transition progressive de l'estimation dite synthétique (méthodes ad hoc dans le cadre général de l'estimation par quotient et par la régression) à l'inférence fondée sur un modèle qui vise à refléter les caractéristiques essentielles de la structure de la population et à fournir des estimations appropriées des paramètres, des mesures de l'incertitude, etc. Il existe aujourd'hui pour les modèles mixtes (s'appuyant, par exemple sur le meilleur prédicteur empirique sans biais, le modèle hiérarchique de Bayes et d'autres méthodes) un ensemble de résultats probants qui s'appuient sur la théorie et les méthodes statistiques générales. Une importance particulière a été accordée à certains modèles, dont :

$$\text{Fay-Herriot :} \quad \theta_i = g(\bar{Y}_i) = z_i^T \beta + v_i, \quad i = 1, \dots, m$$

$$\text{Ordonnée à l'origine aléatoire :} \quad y_{ij} = x_{ij}^T \beta + v_i + e_{ij}$$

Deux niveaux :

$$y_{ij} = x_{ij}^T \beta_i + e_{ij} \quad j = 1, \dots, N_i \quad i = 1, \dots, m$$

$$\beta_i = z_i \gamma + v_i$$

Le modèle de Fay-Herriot (Fay et Herriot, 1979) modélise les moyennes des petites régions (ou les fonctions de ces moyennes) de façon directe grâce à l'utilisation de covariables z_i au niveau des petites régions pour tenir compte de certaines variations inter-régions.

Le modèle avec ordonnée à l'origine aléatoire utilise des variables auxiliaires au niveau de l'unité x_{ij} et tient compte d'autres variations entre domaines en permettant que la valeur de l'ordonnée à l'origine de la régression soit aléatoire. Après l'estimation fondée sur le modèle, les moyennes des petites régions sont essentiellement calculées par agrégation des valeurs prévues pour chaque région et ajout de la valeur prévue de l'ordonnée aléatoire pour la petite région en question.

Le modèle à deux niveaux pousse ces idées une étape plus loin en utilisant des variables auxiliaires au niveau de l'unité x_{ij} , et en permettant que tous les paramètres de régression (et non simplement l'ordonnée à l'origine) soient aléatoires et qu'une partie de la variation entre régions soit expliquée par les variables auxiliaires z_i au niveau de la région. Les termes aléatoires de deuxième niveau (région) peuvent être corrélés.

Il est possible de rendre ces modèles plus généraux et il faut spécifier les structures exactes de l'erreur qui peuvent inclure les structures hétéroscédastiques de la variance et de la covariance, au besoin. En outre, les modèles peuvent être généralisés (p. ex. pour modéliser des variables binaires discrètes). Rao (2003) fait un excellent compte rendu des modèles et des méthodes d'estimation connexes.

L'examen de cette approche en regard des besoins stratégiques du BNS permet de dégager un certain nombre de problèmes qui demanderont un développement méthodologique plus approfondi :

- L'utilisation d'une méthode fondée sur un modèle signifie que la justification habituelle de la qualité des estimations résultantes, à savoir qu'elles ne reposent essentiellement sur aucune hypothèse et sont par conséquent sans biais par rapport à la loi aléatoire, ne tient plus. Le choix de la formulation du modèle, de la structure de l'erreur et des variables auxiliaires, qui est laissé à la discrétion du statisticien officiel chargé de l'analyse, a une incidence sur les estimations résultantes. Donc, le maintien de la confiance du public dans les estimations et la défense contre les critiques externes des résultats dépendront du professionnalisme de l'analyste responsable et, de façon plus générale, de l'estime en laquelle est tenu l'organisme. Dans ce contexte, d'importantes questions se posent quant à l'information à inclure dans le rapport de méthodologie destiné à appuyer les estimations et aux moyens de démontrer la qualité de ces dernières et de présenter les résultats de l'évaluation.

- La richesse même du choix de modèles disponible dans le cadre des modèles mixtes rend nécessaire la poursuite de la recherche portant sur le choix du modèle et les diagnostics d'ajustement. Ces travaux doivent inclure le dépistage des valeurs extrêmes aux niveaux des petites régions et des points de données individuels. Comme il existe fréquemment une interaction entre la spécification de la partie structurelle du modèle et les composantes de l'erreur, il faut que les méthodes de sélection du modèle et les diagnostics en tiennent compte. Qui plus est, l'ajustement du modèle consiste souvent en un ajustement global aux données dans leur ensemble; mais, puisque le produit final sera une estimation pour chaque petit domaine, il est évident que des diagnostics permettant de repérer ceux qui ne semblent pas se conformer au modèle global seront nécessaires.
- Comme une grande partie des statistiques officielles sont des données discrètes, il faudra poursuivre les travaux sur les modèles linéaires mixtes généralisés.
- En pratique, les petites régions sont souvent de tailles différentes, si bien que, dans certains cas, il est possible de recourir à des estimations directes basées sur des données d'enquête (p. ex., une estimation directe pourrait convenir pour l'Ontario, mais non pour l'Île-du-Prince-Édouard). Il faudra donc examiner plus en profondeur la mesure dans laquelle les utilisateurs jugent des estimations hybrides acceptables.
- Enfin, l'orientation générale qui a fait augmenter la demande d'estimations pour de petites régions veut aussi que ces estimations se prêtent à des comparaisons entre régions. Il faudra donc pousser plus loin l'examen des propriétés statistiques de ce genre de comparaisons quand les estimations sont fondées sur des méthodes à modèle mixte.

Dans le cas de nombreuses applications, il est de plus en plus généralement admis que le modèle de Fay-Heriot ou le modèle à ordonnée à l'origine aléatoire reflète correctement la structure de la population et devrait donc être utilisé. Il en est sans doute ainsi parce que les moyennes des petites régions sont essentiellement calculées d'après des prédictions à la moyenne de n'importe quelle variable auxiliaire au niveau de la petite région, et que, par conséquent, la variation des pentes de petite région ne contribue pas à la prédiction. Nous considérons un modèle à deux niveaux simples ne comptant qu'une seule variable auxiliaire :

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + v_{0i} + v_{1i} x_{ij} + e_{ij}$$

$$y_{ij} = \beta_0^* + \beta_1 (x_{ij} - \bar{X}_i) + v_{0i} + v_{1i} (x_{ij} - \bar{X}_i) + e_{ij}$$

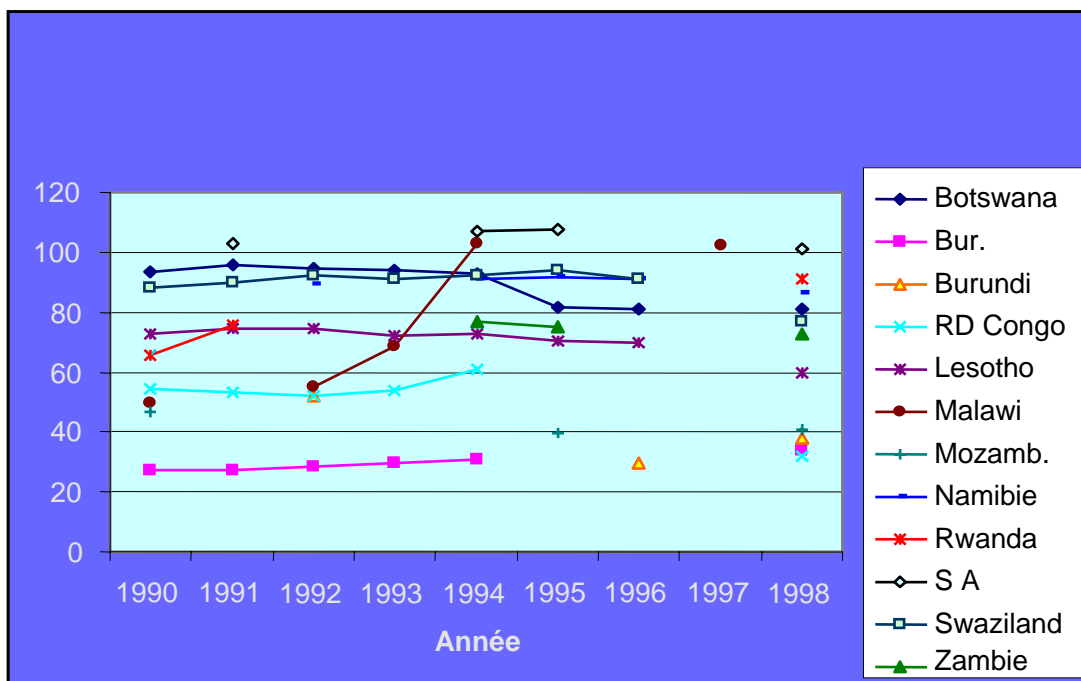
$$\hat{Y}_i = \hat{\beta}_0^* + \hat{\beta}_1 (\bar{x}_i - \bar{X}_i) + \hat{v}_{0i} + \hat{v}_{1i} (\bar{x}_i - \bar{X}_i) \equiv \hat{\beta}_0^* + \hat{v}_{0i}$$

Les pentes aléatoires ne contribuent pas à la moyenne prévue de la petite région, puisque la prédiction est faite pour $\bar{x}_i = \bar{X}_i$.

Cependant, l'un des problèmes de l'estimation de petites régions est de trouver des variables auxiliaires suffisamment informatives pour la prédiction. Une solution évidente consiste à utiliser des données chronologiques provenant des petits domaines de sorte que le temps fasse partie intégrale de la formulation. Dans de nombreuses situations pratiques, la prédiction est faite pour la période la plus récente et l'importance croissante accordée aux « enquêtes continues sur la population » ou aux prétendus « recensements continus » renforcera ce besoin d'estimations pour la « dernière période ». Dans ce cas, la moyenne des variables auxiliaires sera calculée pour la période moyenne sur laquelle les données sont disponibles, tandis que la prédiction sera faite pour la période la plus récente. Dans ces conditions, les pentes variables de la petite région contribuent directement à la prédiction et les arguments en faveur du modèle de Fay-Heriot ou du modèle à ordonnée aléatoire ne sont plus aussi évidents.

Surveillance des indicateurs internationaux de développement

La même question se pose dans le contexte fort différent de la surveillance de la réalisation des objectifs de développement du millénaire dans tous les pays du monde (CSONU, 2004). Ici, l'ensemble de données parfait pour chaque indicateur statistique serait une série chronologique pour chaque pays avec des valeurs de l'indicateur pour chaque année. En pratique, de nombreux pays ne mesurent pas chaque indicateur chaque année et, en fait, les séries chronologiques sont souvent fort incomplètes. La figure 3.1 illustre la situation pour une série de pays d'Afrique subsaharienne pour l'indicateur mesurant le taux de scolarisation au primaire.



Source des données : UNESCO

Figure 3.1 : Taux de scolarisation au primaire dans certains pays d'Afrique

Les séries chronologiques relativement complètes sont rares et pour certains pays, il n'existe qu'une ou deux observations. Les pays sont comparables aux petites régions dont il a été question plus haut et il faut prédire la valeur pour le pays pour une période particulière (souvent la plus récente) et, donc, faire une estimation de l'indicateur pour la région dans son ensemble. Les problèmes qui se posent sont décrits plus en détail dans Holt (2003).

Statistiques de quartier

La demande croissante de « statistiques de quartier » dans un certain nombre de pays peut être considérée comme un autre exemple de l'augmentation de la demande de statistiques pour de petites régions à laquelle doivent répondre les BNS. Toutefois, dans ce cas, il faut aussi résoudre un certain nombre de questions méthodologiques supplémentaires. Jusqu'à présent, la façon dont les BNS ont répondu à cette demande consiste à regrouper l'information statistique sur chaque petit domaine (quartier) à partir de diverses sources, dont le ou les recensements, diverses sources administratives et, dans certains cas, des estimations fondées sur des données d'enquête ou une combinaison de données d'enquête et de données administratives. L'« ensemble de données » résultant est un amalgame d'un éventail d'« objets statistiques » qui peuvent inclure une série de totalisations croisées (calculées d'après les mêmes sources ou des sources différentes). Du point de vue du producteur de statistiques, la méthodologie doit maintenant permettre de produire non seulement des estimations pour petites régions, mais aussi des estimations multidimensionnelles (p. ex., une totalisation croisée) pour chaque petite région. Les besoins de développement méthodologique énumérés plus haut doivent donc être étendus à ce cas.

Cependant, il est juste de se demander si la responsabilité du BNS se limite simplement à produire une série d'estimations reliées ou non reliées pour chaque petite région (les « objets statistiques ») ou si elle va plus loin. Pour de nombreux utilisateurs, les « objets statistiques » sont vraisemblablement un produit intermédiaire qui entrera dans d'autres analyses répondant à leurs besoins réels. Ainsi, les administrations publiques peuvent entrer les valeurs dans des formules de financement ou dans des modèles sur lesquels s'appuie l'élaboration des programmes sociaux. Les administrations publiques et les autorités locales pourraient vouloir suivre certaines mesures de la privation sociale au cours du temps pour évaluer l'effet des programmes sociaux. D'autres pourraient avoir besoin d'« objets statistiques » (p. ex. des totalisations croisées ou des mesures de l'association entre variables) qui ne figurent pas

dans l'ensemble de statistiques de quartier. Ces utilisations soulèvent toute une série de questions méthodologiques concernant les propriétés statistiques de ces utilisations principales de l'ensemble de données qui vont au-delà de la simple détermination des propriétés de chaque « objets statistiques ». Quel genre de soutien le BNS devrait-il offrir pour ce genre d'utilisation? Doit-il étudier les propriétés méthodologiques des utilisations principales (exercice qui, souvent, ne peut être fait que par le BNS, parce qu'il faut avoir accès aux données sous-jacentes) et fournir l'information, l'orientation et les conseils appropriés?

Considérons, par exemple, le cas assez simple de l'estimation de la covariance entre deux variables pour lesquelles aucune information sur la loi conjointe n'est fournie dans l'ensemble de données de quartier. En effet, l'information sur les deux variables provient de sources distinctes. Cette situation aboutit au problème du sophisme écologique (Steel, Holt et Tranmer, 1996) et nous considérons un modèle simple des composantes de la variance pour chacune des deux variables :

$$y_{pij} = \mu_p + v_{pi} + \varepsilon_{pij}, \quad p = 1, 2; \quad i = 1, \dots, m; \quad j = 1, \dots, N_i$$

$$V(v_{pi}) = \sigma_{vp}^2; \quad V(\varepsilon_{pij}) = \sigma_{ep}^2$$

$$Cov(v_{1i}, v_{2i'}) = \sigma_{v12}, \quad i = i'$$

$$Cov(\varepsilon_{1ij}, \varepsilon_{2i'j'}) = \sigma_{\varepsilon12}, \quad i = i', \quad j = j'$$

$$Cov(y_{1ij}, y_{2i'j'}) = \begin{cases} \sigma_{v12} + \sigma_{\varepsilon12} = \sigma_{T12} & i = i', \quad j = j' \\ \sigma_{v12} & i = i' \\ 0 & \text{autrement} \end{cases}$$

Si l'on dispose des variances et des covariances d'échantillon fondées sur les observations conjointes des deux variables, on peut alors obtenir des estimations essentiellement sans biais. Cependant, si les estimations pour les deux variables proviennent de sources différentes, les estimations habituelles ne sont pas disponibles. En particulier, si l'on dispose uniquement des moyennes de petite région pour les deux variables, on pourrait utiliser :

$$\bar{s}_{12} = \frac{1}{m} \sum_i w_i (\bar{y}_{1i} - \bar{y}_1)(\bar{y}_{2i} - \bar{y}_2) \quad w_i = n_{1i} n_{2i} / n_{12i} .$$

On peut montrer que :

$$E(\bar{s}_{12}) = \sigma_{T12} + (\bar{w}^* - 1)\sigma_{v12}, \quad \bar{w}^* = \bar{w} \left(1 - \frac{C_w^2}{m} \right).$$

Donc, l'estimation est biaisée pour σ_{T12} et le biais dépend du degré de chevauchement entre les échantillons sur lesquels les deux moyennes distinctes sont fondées. Dans le cas extrême où il n'existe aucun chevauchement, l'estimation est fondée sur la composante inter-domaines de la covariance et n'inclut pas la composante inter-unités, intra-domaine. L'importance du biais augmente à mesure que le chevauchement diminue.

Une approche adoptée couramment pour contourner ce problème (qui est implicite dans ce qu'il est convenu d'appeler l'appariement statistique) consiste à tenir compte de la loi conjointe régissant les deux variables et certaines variables auxiliaires disponibles pour chacune. En fait, si les deux variables sont présentées dans l'ensemble de données de quartier sous forme de totalisations croisées en fonction, par exemple, du groupe âge-sexe,

elles peuvent être utilisées comme variables auxiliaires. Nous supposons qu'un modèle linéaire relie chaque variable d'intérêt aux variables auxiliaires :

$$y_{pij} = \mu_{p.z} + z_{ij}^T \beta_{pz} + v_{pi} + \varepsilon_{pij}.$$

Steel et coll. (1996) montrent que :

$$E(\bar{s}_{12} | z) = \sigma_{T12} + \beta_{1z}^T (\bar{s}_{z12} - \Sigma_{zz}) \beta_{2z} + (\bar{w}_i^* - 1) \sigma_{12.z}$$

et, par conséquent,

$$\hat{\sigma}_{T12} = \bar{s}_{12} - \beta_{1z}^T (\bar{s}_{12z} - \Sigma_{zz}) \beta_{2z}$$

avec le biais $(\bar{w}^* - 1) \sigma_{v12.z}$.

L'objectif est de trouver les variables auxiliaires Z permettant d'ajuster l'estimation de sorte que le biais résiduel soit réduit au minimum. Cette façon de procéder reflète la composante résiduelle de covariance après avoir tenu compte des variables auxiliaires. Il s'agit essentiellement de l'approche de fusion de données/appariement de données et d'identification. Le problème tient au fait que le biais résiduel sera inconnu (à moins que le BNS ne dispose de données au niveau individuel, mais ne les publie pas pour des raisons de confidentialité). Si le chevauchement entre les sources de données sur les deux variables est faible, même une petite composante de covariance résiduelle sera gonflée par le multiplicateur basé sur la pondération et pourrait être très importante. Si l'une des sources est un recensement (ou une base de données administratives dont la couverture est effectivement complète) et que l'autre est une enquête, le chevauchement entre les deux sources sera fondé sur la taille de l'échantillon de l'enquête pour chaque domaine, puisque $n_{1i}n_{2i} / n_{12i} = n_{1i}$ s'il s'agit de la taille de l'échantillon de l'enquête.

Il convient de souligner que, pour obtenir cet estimateur corrigé, les utilisateurs ont besoin d'une estimation de la matrice des covariances entre les variables auxiliaires Σ_{zz} et que le BNS devra fournir cette estimation en tant qu'élément d'information supplémentaire si elle ne peut être calculée à partir de l'ensemble de statistiques de quartier.

L'objectif de la présente section n'est pas de recommander une solution particulière pour un problème spécifique, mais d'illustrer le fait que les propriétés statistiques des utilisations qui pourraient être faites de l'ensemble de données peuvent dépendre de l'accès aux données sous-jacentes que le BNS est peut-être le seul à posséder et obliger ce dernier à fournir des informations supplémentaires. Ce point demande assurément l'étude méthodologique des utilisations des données qui étendent le rôle du BNS au-delà de la simple production de statistiques pour de petites régions et de leur diffusion.

5. UTILISATION DE DONNÉES ADMINISTRATIVES

Un autre genre de besoin stratégique des BNS tient à l'utilisation croissante de données administratives à des fins statistiques. Cette situation est due en partie à la demande de statistiques pour de petites régions (pour lesquelles les grandes bases de données administratives pourraient permettre la désagrégation au niveau de la petite région que ne permettraient pas les données d'enquête) et en partie à la nécessité de réduire les coûts et le fardeau de réponse. Elle pourrait aussi résulter en partie du fait que certaines sources de données administratives fournissent une couverture plus complète que les enquêtes pour lesquelles on ne connaît pas l'importance du biais de non-réponse. Les contre-arguments habituels sont bien connus, à savoir que :

- souvent les données administratives ne mesurent pas exactement le concept requis;
- souvent la couverture de la source administrative (et la mise à jour pour tenir compte des nouvelles unités et des unités disparues, par exemple) n'est pas d'aussi bonne qualité qu'on le souhaiterait;

- la source administrative n'est ni conçue ni exploitée à des fins statistiques et que cette situation peut avoir une incidence sur la façon dont les données sont utilisées;
- la source administrative n'offre pas autant de variables connexes que ne le ferait un questionnaire d'enquête et que cette situation influe sur la profondeur de toute analyse possible et sur la convergence des estimations avec celles calculées d'après d'autres sources.

Il est possible d'utiliser les sources administratives seules ou, mieux encore, conjuguées à des données d'enquête afin de réduire certaines difficultés susmentionnées. La source administrative peut être utilisée comme base de sondage ou comme source de variables auxiliaires dans le cas de l'estimation par le quotient ou par la régression. Les modes précis d'utilisation dépendront de la possibilité ou non d'apparier les enregistrements individuels provenant des deux sources.

L'usage des données administratives comme base de sondage ou comme source de variables auxiliaires pour l'estimation, que l'on pourrait qualifier d'usage standard, est bien établi dans le cadre habituel d'évaluation de la qualité en statistique officielle. Il existe toutefois des problèmes de couverture et de qualité (en particulier, le fait que nombre de sources administratives présentent des lacunes qui sont plus prononcées pour certains segments de la société que pour d'autres). Ainsi, très souvent, les dossiers fiscaux contiennent systématiquement des lacunes si l'on considère les personnes faiblement rémunérées, ce qui a un effet différentiel sur l'utilisation de cette source de données comme base de sondage, comme source de variables auxiliaires ou comme source de données primaires aux fins de la production de statistiques.

Cependant, j'argumenterai brièvement que le cadre de surveillance de la qualité est nettement moins bien établi pour les données administratives que pour les données d'enquête et qu'il s'agit là d'une lacune significative. Les statisticiens officiels ont passé plus de 40 ans à étudier et à essayer de comprendre l'interaction entre les collecteurs de données et les répondants, ainsi que des aspects tels que l'effet de l'énoncé de la question et de la conception du questionnaire sur les réponses fournies. Plus récemment, la collecte de données assistée par ordinateur a fourni, en tant que sous-produit du processus de collecte, des renseignements supplémentaires qui nous permettent de mieux comprendre le processus proprement dit.

L'échange d'information entre les membres du public et tout système administratif (qui sera la source de données pour un usage statistique ultérieur) est également une forme d'interaction. Contrairement au processus d'enquête, où nous essayons de mesurer les réponses sans influencer sur le comportement réel des répondants, les systèmes administratifs sont souvent conçus pour faire exactement cela. Par exemple, les règles d'admissibilité à un avantage social sont souvent établies en vue d'encourager certaines formes de comportement chez les demandeurs. Alors que les statisticiens officiels essaient de formuler des questions dont l'effet sur le comportement est neutre, les systèmes administratifs sont parfois conçus avec l'intention opposée. Si les statisticiens officiels doivent utiliser ce genre de sources de données à des fins statistiques, alors je soutiens que l'établissement d'un cadre de surveillance de la qualité des données, comparable à celui mis en place pour la méthodologie d'enquête, pourrait être bénéfique. Au minimum, il constituerait un cadre pour l'évaluation de la qualité statistique et, au mieux, il pourrait permettre de mieux comprendre comment concilier les besoins administratifs et statistiques. Si nous voulons offrir aux utilisateurs des statistiques tirées de sources administratives, nous devons mettre en place un meilleur cadre de la qualité dans lequel présenter les résultats.

Erreur de mesure

Un problème de qualité qui se pose aussi bien dans le cas des données d'enquête qu'administratives et auquel on n'a pas accordé suffisamment d'attention est celui de l'erreur de mesure. Fuller (1995) souligne qu'une des hypothèses courantes de modélisation de l'erreur de mesure est que la variable observée x est reliée à la valeur réelle X au moyen d'un terme d'erreur u que l'on suppose dépourvu de biais :

$$x = X + u \qquad E(u) = 0 \text{ et } V(u) = \sigma_u^2.$$

Il fait remarquer que les estimateurs utilisés habituellement pour estimer les moyennes et les variances (par exemple au moyen d'échantillons aléatoires simples) restent sans biais pour la variable observée et qu'on pourrait par conséquent conclure que l'erreur de mesure ne pose pas de problème. Il poursuit que, cependant, même pour

l'estimation de la proportion de la population située au-dessus ou en-dessous d'un seuil particulier, les estimations obtenues ne sont, en général, pas sans biais. Le diagramme schématique simple présenté à la figure 4.1 illustre ce point :

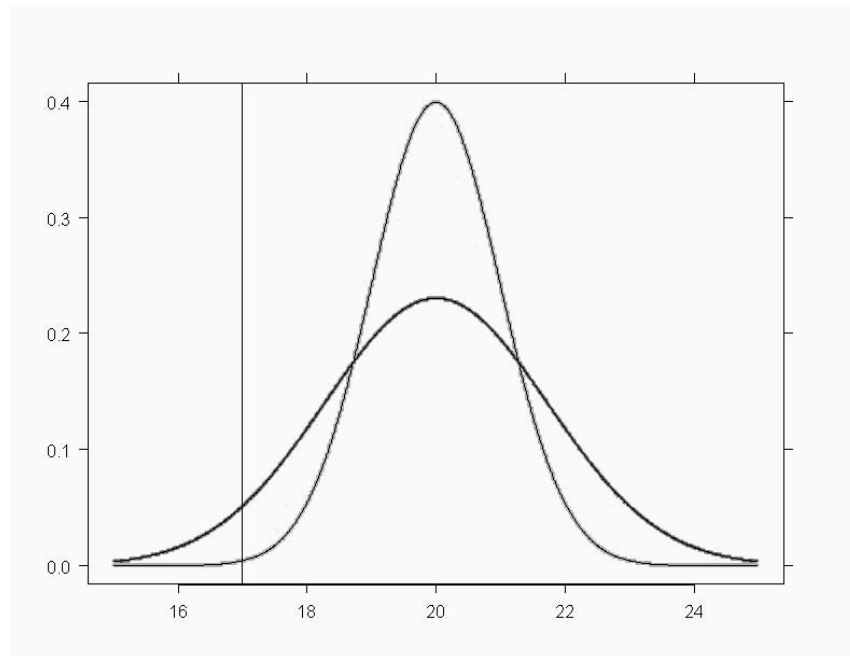


Figure 4.1 : Effets de l'erreur de mesure

La courbe la plus étroite représente la distribution des valeurs réelles et la courbe plus étalée, celle des observations (avec erreur de mesure). L'aire sous la courbe la plus étalée est plus grande pour la proportion de la distribution située en-dessous de tout seuil particulier. Donc, l'erreur de mesure introduit un biais dans l'estimation de la proportion située sous le seuil.

Normalement, ce biais n'est pas facilement observable, mais Skinner et coll. (2002) donnent un exemple fondé sur l'analyse de données administratives où cette distorsion a directement rapport à la question. À la fin des années 1990, le gouvernement du Royaume-Uni a introduit un salaire minimal qui s'applique à la grande majorité de la population active (à quelques exceptions près). Le cas concerne l'estimation de la proportion de personnes gagnant moins que le salaire minimal d'après l'Enquête sur la population active du Royaume-Uni. Il est clair que le calcul de ces estimations joue un rôle important dans la surveillance de l'application de la politique.

Les sources de données sur la rémunération horaire disponibles à des fins statistiques sont une mesure directe du taux salarial horaire provenant d'une source administrative et une mesure dérivée provenant de l'EPA obtenue en divisant le revenu par le nombre habituel d'heures travaillées. La première source présente des problèmes de couverture, particulièrement en ce qui concerne les personnes faiblement rémunérées et la seconde est sujette à une erreur de mesure, en partie parce que le dénominateur est une mesure du nombre habituel d'heures travaillées plutôt que du nombre particulier d'heures travaillées durant la période à laquelle se rapporte le numérateur (rémunération). Une autre mesure directe de la rémunération horaire a été obtenue à partir d'un sous-ensemble de répondants de l'EPA en ajoutant des questions appropriées au questionnaire. Nous nous concentrons ici sur la comparaison entre les mesures dérivée et directe fondées sur les données de l'EPA.

La figure 4.2 (tirée de Skinner et coll., 2002) montre la distribution conjointe des mesures directes ainsi que dérivées.

Cette figure montre clairement l'effet de l'erreur de mesure sur la variable dérivée (x) où la distribution traverse librement la droite du salaire minimal national et où un nombre significatif de répondants sont payés à un taux

horaire inférieur à ce minimum. Comparativement, la distribution de la mesure directe (y) présente une limite beaucoup plus nette à la valeur du salaire minimal.

L'exemple illustre un cas où l'effet de l'erreur de mesure est directement en rapport avec la politique évaluée, puisque la mesure dérivée (avec une erreur de mesure importante) implique qu'un grand nombre de personnes gagnent moins que le salaire minimal et, donc, que l'esprit de la politique n'est pas respecté. La mesure directe corrige considérablement cette impression et brosse un meilleur tableau de la situation réelle. Dans le cas d'autres utilisations de données sujettes à une erreur de mesure, l'effet pourrait être tout aussi important, mais ne pas être apparent, parce qu'il n'existe aucune donnée de référence externes permettant de vérifier la validité des résultats.

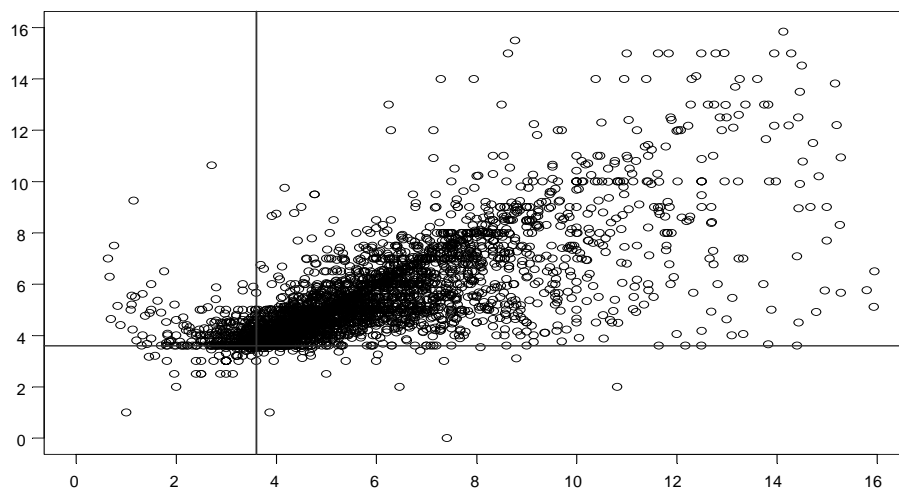


Figure 4.2 : Distribution conjointe des mesures directes (y) et dérivées (x) de la rémunération horaire (£ par heure) provenant de l'Enquête sur la population active du Royaume-Uni (La valeur du salaire minimal national est représentée par les droites verticale et horizontale.)

Source : Skinner et coll. (2002)

6. MESURE DU RENDEMENT

La dernière grande section de l'article traite de la demande comparative récente de mesures statistiques de rendement à un niveau désagrégé, habituellement, celui d'une unité fournissant un service public (voir Bird et coll., 2004). Il pourrait s'agir, par exemple, de la proportion de sortants ayant dépassé un niveau spécifié de rendement scolaire (tel que mesuré par les résultats des examens) selon l'école, de la longueur des listes d'attente pour une intervention chirurgicale non urgente selon l'hôpital ou du taux d'affaires criminelles tirées au clair selon l'autorité policière. Il existe aussi certains exemples à des niveaux d'agrégation plus faibles, comme les taux de mortalité par équipe chirurgicale ou par chirurgien. Souvent, les valeurs des indicateurs sont présentées sous forme de « tableaux de ligue » ou de mesures de rendement par nombre d'étoiles (p. ex. hôpitaux 3*, 2* et 1*).

La motivation de cette demande est le besoin de fournir des mesures de rendement au niveau de l'unité pour deux raisons : pour aider les membres du public (en tant que consommateurs de services publics) à choisir l'unité à utiliser et pour accroître l'obligation qu'ont les ministres, ainsi que les employés et la direction des unités fournisseuses de services publics de rendre des comptes aux membres du public. En ce qui concerne la reddition de comptes, il est fréquent que des cibles précises de rendement soient établies afin de permettre aux membres du public d'évaluer le rendement (p. ex. 75 % d'élèves du secondaire atteindront un niveau précis de rendement scolaire). En pratique, trois éléments sont requis, à savoir le choix du concept à évaluer, l'élaboration et la surveillance d'une mesure statistique appropriée et, enfin, l'établissement d'une cible de rendement fondée sur la mesure choisie.

Malheureusement, ces trois aspects distincts sont souvent confondus, et il arrive que les politiciens ou les agents responsables des politiques formulent le concept en fonction d'une mesure particulière et établissent une cible sans tenir compte de l'utilité statistique, voire même de la faisabilité, de la mesure choisie.

Cette utilisation des indicateurs statistiques soulève plusieurs questions de méthodologie qui dépassent le cadre méthodologique habituel de la statistique officielle. La réponse à ces questions nécessite un développement méthodologique plus approfondi ou, si les travaux de développement ont été réalisés dans d'autres circonstances, l'intégration des méthodes dans la culture du BNS.

Conception statistique

Du point de vue du BNS, le choix d'un indicateur statistique comporte deux aspects importants. Premièrement, il importe que le concept choisi à des fins de reddition de comptes soit accepté par un groupe qui ne se limite pas aux individus responsables de la prestation du service public. Si le concept est perçu comme étant un « choix politique », le BNS risque d'être considéré comme appuyant une mesure dont le mobile est politique, ce qui pourrait entraîner la mise en question de son indépendance statistique et de son intégrité. Deuxièmement, même si le concept est valable, il existe un compromis inévitable entre ce qu'il est conceptuellement souhaitable et ce qu'il est statistiquement possible de mesurer. Troisièmement, le choix d'une cible de rendement est une décision politique, quoiqu'une analyse statistique pourrait fort bien l'éclairer.

Il existe des problèmes de conception que les statisticiens officiels ont l'habitude de résoudre, comme savoir s'il est préférable de recueillir de l'information auprès de la totalité des unités prodiguant les services ou si l'échantillonnage des cas/sujets serait plus efficace. Par exemple, dans le cas des tests aléatoires obligatoires de dépistage de substances illicites dans le sang chez les prisonniers, le processus de collecte des données sous-entend que les coûts sont en grande partie proportionnels au nombre d'échantillons prélevés dans chaque établissement, alors que dans le cas de l'information sur le rendement scolaire, des données peuvent être recueillies d'après les dossiers scolaires des élèves d'une école en n'occasionnant essentiellement aucun coût supplémentaire.

Normalement, on devrait examiner la grandeur d'un effet possible d'une politique et calculer la taille d'échantillon nécessaire pour déceler cet effet avec la puissance requise au niveau de l'unité déclarante. En revanche, toute cible choisie par les agents responsables des politiques ou par les ministres risque de refléter des aspirations plutôt que le niveau de l'effet qu'il serait important de mesurer ou raisonnable d'atteindre. Choisir les tailles d'échantillon permettant d'obtenir une puissance suffisante pour déceler des variations ou des tendances significatives pourraient aller au-delà du choix d'une cible dans l'espoir d'atteindre un but.

Fait plus important, certaines petites unités déclarantes pourraient simplement ne pas générer le nombre d'observations requises, même dans le cas d'un recensement, pour obtenir une puissance suffisante pour déceler les variations attendues (voir plus loin).

Qualité et vérification des données

La façon dont sont utilisés les indicateurs statistiques peut avoir des conséquences importantes pour les personnes faisant l'objet de la surveillance. Des résultats insatisfaisants risquent de donner lieu à des sanctions personnelles ou organisationnelles et des résultats satisfaisants peuvent se traduire par une récompense pour les gestionnaires (rémunération reliée au rendement) et pour l'organisme (plus grande autonomie financière ou de gestion). Dans certains cas, des résultats insatisfaisants aboutissent à l'imposition d'un changement de direction ou de structure organisationnelle, sapent le moral des employés de l'unité qui prodigue les services et rendent plus difficile le recrutement d'employés de haut calibre. Par conséquent, les employés des unités de service ont des motifs importants de s'assurer que l'indicateur statistique soit aussi satisfaisant que possible. Dans la mesure où ils donnent lieu à une modification véritable de l'efficacité opérationnelle de l'unité, ces motifs font partie de la logique qui sous-tend la politique, mais ils peuvent avoir un effet pervers entraînant des comportements susceptibles d'altérer l'indicateur statistique sans modifier le rendement sous-jacent.

Dans les conditions extrêmes, l'effet pervers peut pousser à la falsification des données ou, si le codage des données comporte tout élément discrétionnaire, à des pratiques d'enregistrement qui influencent la valeur de l'indicateur. Par exemple, dans le cas d'un indicateur mesurant la violence des prisonniers à l'égard d'autres prisonniers, il pourrait

être nécessaire que le personnel de la prison mène une enquête au sujet de chaque prétendu événement afin de déterminer s'il a effectivement eu lieu. Il est probable que le nombre d'événements enregistrés sera d'autant plus élevé que l'enquête concernant chaque événement éventuel est rigoureuse. Par conséquent, le personnel de la prison pourrait être enclin à qualifier certains événements de « comportement généralement indiscipliné » au lieu de les considérer comme des cas éventuels de violence de prisonniers à l'égard d'autres prisonniers.

Du point de vue de la conception des mesures, l'implication méthodologique est que les définitions et les critères de classification des données doivent être suffisamment précis pour éviter les effets pervers sur le comportement et que les procédures de vérification de la qualité des données doivent être sensibles à ces effets éventuels.

Un deuxième exemple d'effet pervers qui n'aboutit pas à la falsification des données, mais peut aller à l'encontre de l'amélioration du service que la surveillance du rendement a pour but de favoriser est donné par Dranove et coll. (2002). Après la publication du rendement, tel que mesuré par les résultats observés chez les patients, d'une liste nominative de chirurgiens cardiaques new-yorkais, une évaluation a montré que le comportement d'aversion au risque subséquent des chirurgiens s'est concrétisé par une réticence à opérer les patients les plus gravement malades (c'est-à-dire ceux dont le pronostic était le plus mauvais) et un accroissement compensatoire du nombre d'interventions médicales chez les patients moins gravement atteints. L'effet global de ce comportement d'aversion au risque a été une détérioration des résultats observés chez les patients.

Analyse et interprétation

L'analyse et l'interprétation des données utilisées pour la surveillance du rendement posent d'importantes questions d'ordre méthodologique. Il faut savoir comment tenir compte de la composition de la clientèle des unités de prestation de services et du contexte dans lequel elles fonctionnent, et comment interpréter les mesures en présence de la variation inhérente des données.

Prise en compte de la composition de la clientèle

L'objectif fondamental de l'évaluation du rendement est de surveiller le rendement de l'unité fournissant les services. Considérons, par exemple, la surveillance du rendement des écoles d'après la proportion de sortants atteignant un standard donné. Il serait raisonnable de penser qu'une école desservant un quartier déshérité dont les élèves proviennent de collectivités où diverses langues sont parlées quotidiennement devra relever des défis en matière d'éducation nettement plus importants qu'une école desservant un quartier plus favorisé. Ou bien, si le critère de rendement scolaire requis pour l'entrée dans chaque école était mesuré d'après une évaluation nationale, on pourrait s'attendre à ce que les écoles exigeant de plus grandes capacités à l'entrée produisent subséquemment les sortants ayant les capacités les plus grandes.

Il semble donc que la mesure du rendement des écoles doit tenir compte des variations qui influent sur la comparabilité des données afin d'obtenir des mesures de rendement que nous conviendrons d'appeler « à valeur ajoutée ». La même situation se produit dans le contexte médical où il est essentiel de tenir compte de la composition de la clientèle (par exemple, une mesure de la gravité de la maladie au début du traitement) si l'on veut évaluer l'effet du traitement indépendamment de la gravité des cas.

Soulignons qu'il n'est pas toujours facile de se mettre d'accord sur les sources de variation qu'il convient de corriger. Par exemple, lors de la comparaison d'écoles, faut-il corriger le niveau final de rendement scolaire en tenant compte du niveau d'absentéisme dans l'école? L'argument en faveur de cette correction est que, s'il est élevé, le niveau d'absentéisme aura, en principe, des répercussions sur le niveau de rendement de l'école (correction pour une variable exogène). L'opinion contraire est que le niveau d'absentéisme reflète la mesure dans laquelle l'école suscite l'enthousiasme des élèves et leur intérêt pour l'éducation et reflète donc un élément sur lequel l'école pourrait exercer une influence (pas de contrôle pour une variable endogène).

Notons aussi que, même si tenir compte des variations inhérentes est la bonne solution, il n'est pas toujours facile de mesurer avec précision la variable de contrôle. L'évaluation de la gravité d'une maladie au moment où débute le traitement médical n'est pas nécessairement simple.

En supposant que l'on puisse résoudre ces problèmes et produire une estimation à « valeur ajoutée » du rendement, d'autres problèmes méthodologiques se posent aux statisticiens officiels. L'un des moyens courants de les aborder consiste à utiliser un modèle mixte tel que décrit à la section 4 :

$$y_{ij} = x_{ij}^T \beta_i + e_{ij} \quad j = 1, \dots, N_i \quad i = 1, \dots, m$$

$$\beta_i = z_i \gamma + v_i$$

Dans le cas du rendement des écoles, par exemple, i peut être l'indice représentant les écoles et j , celui représentant les étudiants individuels. Les variables x représentent les covariables de niveau individuel et z , les covariables d'école ou de quartier.

Le plus fréquemment, on choisit un modèle à ordonnée aléatoire, de sorte qu'il n'existe dans le modèle qu'un seul terme aléatoire v associé à l'ordonnée à l'origine.

Dans le cas des modèles de ce genre proposés pour l'estimation sur petits domaines à la section 4, il existe une solution convergente par rapport au plan de sondage, même si elle n'est pas efficace. Dans le cas de la surveillance du rendement, la formulation du modèle fait partie intégrante de l'objectif. La mesure à valeur ajoutée du rendement est la valeur prévue du terme aléatoire v , puisque l'objectif consiste à mesurer le rendement de l'unité conditionnellement aux covariables. Donc, la formulation du modèle n'est pas simplement un moyen commode d'estimer un paramètre de population pour un petit domaine, mais un moyen de déterminer l'objet de l'inférence. La valeur de la mesure du rendement dépendra du choix des covariables aux niveaux de l'élève et de l'école, ainsi que de la façon précise dont les termes sont inclus dans la formulation des composantes aléatoires multiniveaux.

Produire des mesures à valeur ajoutée du rendement à l'intention du public représente un grand pas par rapport au genre de statistiques que produit habituellement un BNS. Le degré de confiance et d'acceptation du public dépendra de la réputation professionnelle du BNS, de la qualité de l'analyse sous-jacente et de la démonstration, grâce à des rapports méthodologiques, de la validité des méthodes de sélection du modèle et de l'utilisation des mesures diagnostiques.

Il convient de souligner que, s'il n'existe qu'un seul terme aléatoire au niveau de l'école, à savoir l'ordonnée aléatoire, l'interprétation du rendement à valeur ajoutée est claire. Cependant, s'il faut inclure des composantes supplémentaires, comme des coefficients aléatoires au niveau de l'école pour les covariables, alors l'interprétation du rendement de l'école devient beaucoup plus difficile à communiquer. Supposons, par exemple, qu'il existe un terme aléatoire associé au sexe de l'élève, autrement dit que le rendement relatif des écoles ne sera pas le même pour les élèves de sexe masculin et féminin. En général, les décideurs s'efforcent d'utiliser une mesure simple, dont la signification est facile à communiquer, du rendement des écoles. La description des résultats, s'il existe des différences de ce genre entre les écoles, compliquera leur présentation.

Variation inhérente

Une autre question d'ordre méthodologique est celle de savoir comment présenter les résultats en tenant compte de la variation inhérente. Trop souvent, nous interprétons des données, présentons des tableaux de ligue ou classons les unités prestataires de services dans la catégorie des unités exemplaires ou de celles qui échouent d'après la valeur la plus récente des données, sans tenir compte comme il convient de l'incertitude inhérente à l'estimation. Les données (p. ex. le taux de d'élucidation des affaires criminelles ou le taux de réussite scolaire) sont traitées comme s'il s'agissait d'une mesure précise du rendement sous-jacent. Par conséquent, on juge les écoles, les chirurgiens ou les autorités policières comme étant exemplaires ou ayant échoué d'après les données les plus récentes.

En pratique, l'incertitude associée à la prédiction fondée sur un modèle mixte du rendement d'une unité peut être considérable et, dans certaines situations, il n'existe aucun remède. Si l'unité d'analyse est une école qui ne produit que 100 sortants annuellement, ce sont les seules données dont on dispose (à moins que l'on évalue le rendement de l'école en regroupant les données sur plusieurs années, ce qui pourrait aller à l'encontre du but recherché par la direction).

Les limites des tableaux de ligue dans lesquels les unités prestataires de services sont classées d'après la mesure du rendement sont bien connues. Goldstein et Spiegelhalter (1996) font un bon compte rendu de l'imprécision.

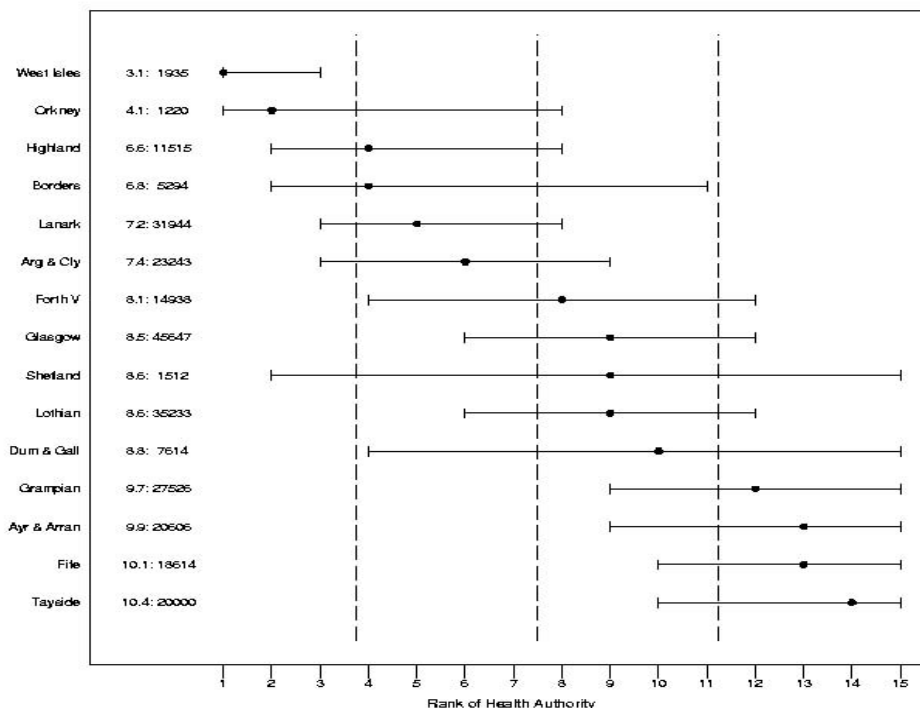


Figure 4.3 : Médianes et intervalles de confiance à 95% pour le classement des autorités sanitaires écossaises en fonction du taux de conception chez les adolescentes, 1990-1992
 Source : Bird et coll. (2004), tiré de Goldstein et Spiegelhalter (1996).

La figure 4.3, qui est tirée de Goldstein et Spiegelhalter (1996), montre la médiane et l'intervalle de confiance à 95 % pour les autorités sanitaires écossaises en fonction du taux de conception chez les adolescentes. Le graphique démontre que seule l'autorité pour laquelle le taux est le plus faible peut être classée avec certitude dans le premier quartile et les quatre autorités affichant les taux les plus élevés, dans la moitié supérieure du classement par ordre. À part cela, l'incertitude est trop importante pour tirer des conclusions catégoriques quant au classement des autres autorités.

En pratique, le changement de classement d'une année à l'autre (en tant que mesure de la variation du rendement) serait une mesure encore plus incertaine.

Trop fréquemment, les tableaux de ligue ou les classements par nombre d'étoiles (3*, 2* et 1*) sont produits et interprétés avec un niveau hypothétique de précision que ne justifie simplement pas le degré d'incertitude de l'inférence. Aider les utilisateurs à comprendre les limites des tableaux et à éviter les surinterprétations est une tâche importante.

Une question connexe est que, souvent, les cibles de rendement sont établies en cascade, sans bien comprendre les conséquences. Supposons, par exemple, que l'on fixe pour cible que 75 % des sortants d'une école devraient atteindre un standard d'apprentissage particulier. Au niveau national, au Royaume-Uni (où l'on compte environ

600 000 élèves par année), pour être certain à 99 % d'atteindre la cible, le système devrait fonctionner avec une proportion un peu plus forte que 75 % d'élèves atteignant la cible afin de tenir compte de toute fluctuation aléatoire qui pourrait survenir durant une année. Pour que l'enseignant d'une classe de 30 élèves soit certain à 99 % d'atteindre la cible, le taux de rendement devrait être de près de 90 %. Si l'enseignant et le pays devaient l'un et l'autre fonctionner avec un taux de réussite de 80 %, la mesure du rendement national excéderait presque systématiquement la cible, alors que l'enseignant courrait chaque année un grand risque de ne pas l'atteindre. Nous notons aussi qu'il n'est pas possible de réduire l'intervalle de confiance pour l'enseignant individuel, puisque tous les élèves de la classe sont évalués. L'incertitude est inhérente aux propriétés des données et à l'inférence.

L'analyse et la présentation des mesures de rendement soulèvent donc diverses questions de nature méthodologique. Il est souvent essentiel de tenir compte de la composition de la population desservie par l'unité qui fournit les services, mais les covariables appropriées ne sont pas toujours disponibles; la formulation précise du modèle, qui demande un jugement professionnel, a des répercussions sur les mesures du rendement; le paramètre d'intérêt fait partie intégrante de la formulation du modèle, et les variations inhérentes aux mesures et la façon dont elles sont interprétées soulèvent des questions importantes. Les bureaux nationaux de la statistique devront résoudre ces problèmes s'ils sont obligés de produire des mesures de rendement.

7. SOMMAIRE ET CONCLUSION

Brièvement, l'article fait valoir que les développements méthodologiques ont de l'importance, non seulement comme moyen d'assurer la qualité et l'acceptabilité des produits statistiques individuels, mais aussi au-delà. De manière systématique, ces développements établissent la réputation professionnelle du BNS, contribuent à l'élaboration de normes internationales et, donc, renforcent la confiance des membres du public dans le BNS et ses produits.

Étant donné le rôle stratégique de la méthodologie au sein d'un BNS, il est soutenu que la planification des développements méthodologiques futurs doit tenir compte des besoins stratégiques de l'organisme. La tâche n'est pas simple, mais l'article essaye d'établir des liens entre les développements stratégiques et les conséquences méthodologiques. Il est essentiel de mettre en place un processus permettant d'établir ce lien s'il l'on veut que les futurs développements méthodologiques répondent aux besoins stratégiques du BNS.

RÉFÉRENCES

- Bird S M, Cox D R, Farewell V, Goldstein H, Holt D, Smith P C (2004) Performance Indicators: Good, Bad and Ugly. *J Roy. Statist. Soc* (to appear)
- Dranove D, Kessler D, McClellan M, Satterthwaite M (2002). Is more information better? The effects of report cards on Health Care Providers. National Bureau of Economic Research, working paper No w8697, Cambridge Ma.
- Fay R E et Herriot R A (1979). Estimation of Income from Small Places: An Application of James-Stein Procedures to Census Data. *J. Amer. Statist. Assoc.* 74, 269-277.
- Fuller W A (1995). Estimation in the Presence of Measurement Error. *International Statistical Review*, 63, 121-141.
- Goldstein H et Spiegelhalter D J (1996). League Tables and their Limitations: Statistical Issues in the Comparison of Statistical Performance (with Discussion). *J. Roy. Statist. Soc. A*, 159, 3, 385-443.
- Holt D (2003). Methodological Issues in the Development and Use of Statistical Indicators for International Comparisons. *Survey Methodology*, 29, 1, 5-18.
- Neyman J (1934). On the Two Different Aspects of the Representative Method: the Method of Stratified Sampling and the Method of Purposive Selection. *J. Roy. Statist. Soc.*, 97, 558-625.

Rao J N K (2003). *Small Area Estimation*. John Wiley, New Jersey.

Skinner C J, Stuttard N, Beissel-Durrant G et Jenkins J (2002). *The Measurement of Low Pay in the UK Labour Force Survey*. *Oxford Bulletin of Economics and Statistics*, 64, Supplement, 653-676.

Steel D G, Holt D et Tranmer M (1996). *Making Unit-Level Inferences from Aggregated Data*. *Survey Methodology*, 22, 1, 3-15.

UNSC (2002) *An Assessment of the Statistical Indicators Derived from United Nations Summit Meetings*. UN Statistical Commission, New York.

UNSD (2004). *Millennium Indicators Data-base*. http://unstats.un.org/unsd/mi/mi_goals.asp