No 11-522-XIF au catalogue

La série des symposiums internationaux de Statistique Canada - Recueil

Symposium 2005 : Défis méthodologiques reliés aux besoins futurs d’information

2005

Statistique Canada Statistics Canada

Canada

Recueil du Symposium 2005 de Statistique Canada
Défis méthodologiques reliés aux besoins futurs d'information

SONDAGE INDIRECT À DEUX PHASES : UNE APPLICATION À L'ESSAI DE QUESTIONNAIRES SUR LE TERRAIN

M.A. Hidiroglou et Pierre Lavallée[1]

RÉSUMÉ

Il existe plusieurs façons d'améliorer la qualité des données. L'une d'entre elles consiste à refondre et à mettre à l'essai les questionnaires des enquêtes permanentes. La refonte et l'essai des questionnaires offrent l'avantage d'améliorer l'exactitude en s'assurant que les questions servent à recueillir les données nécessaires, ainsi que de réduire le fardeau de réponse. Un certain nombre de questionnaires d'enquêtes-entreprises du Bureau de la statistique nationale (*Office for National Statistics*) sont en cours de refonte. Dans le présent exposé, nous allons nous concentrer sur la refonte du questionnaire de l'Enquête annuelle sur les heures et les gains (*Annual Survey of Hours and Earnings (ASHE*)).

MOTS CLÉS :Sondage à deux phases; conception de questionnaire; sondage indirect.

1. INTRODUCTION

L'Enquête annuelle sur les heures et les gains (*Annual Survey of Hours and Earnings, ASHE*) est la principale source de données sur la répartition des gains au Royaume-Uni. Elle est menée annuellement depuis 1970, sans changement notoire. En avril de chaque année, elle mesure les gains d'emploi des travailleurs du Royaume-Uni dans l'ensemble des secteurs économiques. L'ASHE est menée par le Bureau de la statistique nationale (*Office for National Statistics, ONS*) de la Grande-Bretagne et par le Département de l'entreprise, du commerce et de l'investissement (*Department of Enterprise, Trade and Investment*) de l'Irlande du Nord.

La majeure partie de l'échantillon de 1 % des employés est tirée par *Inland Revenue* (le département du Royaume-Uni chargé de percevoir l'impôt sur le revenu) à partir du système PAYE (*Pay As You Earn*), en fonction des deux derniers chiffres du numéro d'assurance nationale de chaque employé. Une petite proportion de l'échantillon est identifiée directement, selon le même critère de sélection, par les employeurs qui retournent l'information électroniquement à l'ONS. Le plan de sondage crée, en fait, un panel d'employés. Depuis quelques années, la taille de l'échantillon est d'environ 240 000 employés. Pour plus de détails concernant l'ASHE, voir Pont (2003).

Les numéros d'assurance nationale sélectionnés sont appariés au Registre interdépartemental des entreprises (*Interdepartmental Business Register, IDBR*) de l'ONS au niveau des entreprises. Les entreprises identifiées doivent alors remplir un questionnaire pour chacun de leurs employés sélectionnés d'après le fichier PAYE de *Inland Revenue*. Les données demandées au sujet des employés comprennent, entre autres, leurs gains, leurs heures travaillées et une description de leur occupation. Dans le présent exposé, on suppose qu'un employé est rattaché à une seule entreprise.

Jusqu'en 2004, on a utilisé un questionnaire de deux pages pour recueillir ces données auprès des entreprises visées. En 2004, on a mis à l'essai un nouveau questionnaire de six pages en vue de l'utiliser en 2005. Ce nouveau questionnaire comprenait un certain nombre de questions supplémentaires et l'on avait amélioré la présentation. Le taux de réponse étant l'une des mesures susceptibles d'indiquer si la nouvelle version du questionnaire constituait ou non une amélioration, on a scindé l'échantillon en deux sous-échantillons pour déterminer s'il y avait un écart significatif entre les taux de réponse aux deux questionnaires. À une partie de l'échantillon, on a envoyé par la poste

[1] M.A. Hidiroglou, Office for National Statistics, Cardiff Road, Newport, Royaume-Uni, NP10 8XG (mike.hidiroglou@ons.gsi.gov.uk); Pierre Lavallée, Division des méthodes d'enquêtes sociales, Statistique Canada, Ottawa (Ontario), Canada, K1A 0T6 (pierre.lavallee@statcan.ca)

5 000 questionnaires refondus (version de six pages) et, au reste de l'échantillon, 235 000 questionnaires de deux pages. L'échantillonnage a respecté la contrainte voulant que les employés sélectionnés d'une même entreprise (ou grappe) reçoivent tous le même questionnaire (celui de deux pages ou celui de six pages). Comme nous le verrons plus loin, cette contrainte complique davantage l'estimation de la variance.

L'exposé est structuré comme suit : dans la section 2, nous décrivons en détail la méthode d'échantillonnage des deux questionnaires, ainsi que la notation employée. Dans la section 3, nous décrivons la variance de la réponse estimée à chaque questionnaire. Pour vérifier si les taux de réponse aux deux types de questionnaire sont différents, on a besoin d'un terme de covariance; nous abordons cet aspect dans la section 4. Enfin, nous résumons nos constatations dans la conclusion.

2. RÉPARTITION DES DEUX QUESTIONNAIRES

Soit U^A, la base de sondage de M employés (*Inland Revenue*) d'après leurs numéros d'assurance nationale. Soit U^B, l'univers correspondant de N entreprises appariées dans l'IDBR. Chaque entreprise de la base de sondage U^B peut être considérée comme une grappe d'employés rattachés aux entreprises de la base de sondage U^A. Un premier échantillon s^A de m employés (éléments) est tiré de U^A par échantillonnage de Bernoulli. Dans ce qui suit, on calcule les variances estimées en supposant que s^A est tiré par sondage aléatoire simple sans remise de U^A. On identifie les employés par l'indice k. L'échantillon s^A est scindé en deux sous-échantillons s_2^A et s_6^A auxquels on envoie respectivement le questionnaire de deux pages et celui de six pages. Ces deux sous-échantillons contiennent respectivement m_2^A et m_6^A employés.

La répartition des employés compris dans s^A aux deux sous-échantillons est effectuée de sorte qu'une entreprise reçoit le même questionnaire, de deux ou de six pages, pour tous ses employés sélectionnés. L'échantillon correspondant de n^B entreprises (grappes) tiré du Registre des entreprises et apparié aux éléments de l'échantillon s^A de numéros d'assurance nationale sélectionnés est représenté par s^B, où $s^B = \left\{ i \mid k \in i,\ i \in U^B, k \in s^A \right\}$. L'échantillon s^B est stratifié en H strates s_h^B composées de n_h^B grappes, $h=1,\ldots, H$: $s^B = \bigcup_{h=1}^{H} s_h^B$. On tire de s_h^B un échantillon aléatoire simple sans remise s_{6h}^B de n_{6h}^B grappes. À chaque élément compris dans les grappes sélectionnées, on envoie par la poste un questionnaire de six pages. Au reste de l'échantillon s_{2h}^B comprenant n_{2h}^B grappes, on envoie par la poste un questionnaire de deux pages. Les éléments correspondants de s^A appariés aux éléments appartenant à s_{2h}^B et s_{6h}^B sont représentés respectivement par s_{2h}^A et s_{6h}^A, où $s_{2h}^A = \left\{ k \mid k \in i, i \in s_{2h}^B, k \in s^A \right\}$ et $s_{6h}^A = \left\{ k \mid k \in i, i \in s_{6h}^B, k \in s^A \right\}$. Il est à noter que $s_h^A = s_{2h}^A \cup s_{6h}^A$, $s_h^B = s_{2h}^B \cup s_{6h}^B$, $m_h = m_{2h} + m_{6h}$ et $n_h^B = n_{2h}^B + n_{6h}^B$, $h=1,\ldots,H$.

On peut dire que le plan de sondage est un plan de sondage indirect à deux phases. Il est indirect, car on sélectionne d'abord les employés dans la base de sondage de *Inland Revenue*, puis on les apparie aux entreprises inscrites dans le Registre des entreprises, auxquelles on demande ensuite des données sur ces employés. Il est à deux phases, car les questionnaires de deux pages ou de six pages constituent des sous-échantillons des entreprises appariées. Le plan de sondage à deux phases permet d'intégrer les expériences dans l'échantillonnage d'enquête.

Après avoir recueilli les données, on vérifie s'il existe un écart significatif entre les taux de réponse aux questionnaires de deux pages et de six pages. Le taux de réponse au questionnaire de deux pages est estimé comme suit : $p_2 = \hat{Y}_2 / \hat{M}_2$, où $\hat{Y}_2$ est la réponse pondérée globale et $\hat{M}_2$ est le nombre pondéré correspondant de questionnaires de deux pages envoyés par la poste. Le nombre total de questionnaires de deux pages envoyés par la poste de la $i^{\text{ième}}$ grappe comprise dans la $h^{\text{ième}}$ strate (s_{2hi}^B) est m_{2hi}. L'ensemble des éléments compris dans s^A qui correspondent à ceux des grappes s_{2hi}^B est représenté par s_{2hi}^A; dans ce qui suit, nous utilisons ces deux sous-ensembles de manière interchangeable. Les totaux estimés $\hat{Y}_2$ et $\hat{M}_2$ sont respectivement

$\hat{Y}_2 = \frac{M}{m} \sum_{h=1}^{H} \frac{n_h^B}{n_{2h}^B} \sum_{i=1}^{n_{2h}^B} \sum_{k \in s_{2hi}^A} y_{2k}$ et $\hat{M}_2 = \frac{M}{m} \sum_{h=1}^{H} \frac{n_h^B}{n_{2h}^B} \sum_{i=1}^{n_{2h}^B} m_{2hi}$. La variable y_{2k} indique l'état de la réponse : elle correspond à un si l'on a répondu au questionnaire, et à zéro dans le cas contraire. On estime de la même manière le taux de réponse au questionnaire de six pages, représenté par p_6.

On vérifie l'écart entre les taux de réponse aux questionnaires de deux pages et de six pages à l'aide du test *t* donné par

$$t = \frac{p_2 - p_6}{\sqrt{\mathrm{v}(p_2) - 2\,\mathrm{cov}(p_2, p_6) + \mathrm{v}(p_6)}} \tag{1}$$

3. ESTIMATION DE LA VARIANCE DE p_2

Dans la présente section, nous nous limitons à l'estimation de la variance du taux de réponse estimé p_2 pour le questionnaire de deux pages, puisqu'on procède de la même façon pour estimer la variance de la proportion estimée pour le questionnaire de six pages. Nous linéarisons d'abord la proportion $p_2 = \hat{Y}_2 / \hat{M}_2$ (estimateur par quotient), où $\hat{Y}_2$ et $\hat{M}_2$ sont respectivement les estimateurs sans biais de $Y_2 = \sum_{k \in U^A} y_k$ et *M*. La variable linéarisée étant définie comme suit : $z_{2,k} = \left(y_{2,k} - p_2\right) / \hat{M}_2$, la variance estimé de p_2 est donnée par $\mathbf{v}(p_2) \doteq \mathbf{v}(\hat{Z}_2)$ où $\hat{Z}_2 = \frac{M}{m} \sum_{h=1}^{H} \frac{n_h^B}{n_{2h}^B} \sum_{i=1}^{n_{2h}^B} \sum_{k \in s_{2hi}^A} z_{2,k}$ et $\mathrm{v}(\bullet)$ représente l'opérateur de variance qui reflète le plan de sondage.

Soient π_{1k} et $\pi_{1k\ell}$ qui représentent les probabilités d'inclusion du premier et du second ordre pour le premier échantillon s^A. De même, soient $\pi_{2i|s^A}$ et $\pi_{2ij|s^A}$ les probabilités d'inclusion conditionnelles du premier et du second ordre pour le deuxième échantillon $s_2^A = \bigcup_{h=1}^{H} s_{2h}^A$, compte tenu de l'échantillon de première phase s^A. Nous utilisons l'estimateur de variance de Sen-Yates-Grundy (SYG) pour le sondage à deux phases, proposé par Hidiroglou et Rao (2003), pour obtenir la variance estimée de $\hat{Z}_2$. Cet estimateur de variance donne moins souvent des estimations négatives de la variance que l'estimateur de Horvitz-Thompson (HT) pour les plans de sondage non fixes. La variance obtenue est toujours non négative pour notre plan de sondage. L'estimateur de variance SYG à deux phases est donné par

$$\mathrm{v}_{\mathrm{SYG}}\left(\hat{Z}_2\right) = \mathrm{v}_{2,SYG}\left(\hat{Z}_1\right) + \mathrm{v}_{\mathrm{SYG}}\left(\hat{Z}_2 \middle| s^A\right) \tag{2}$$

où $\hat{Z}_1 = E\left(\hat{Z}_2 \middle| s^A\right) = \sum_{s^A} z_k / \pi_{1k}$, $\mathrm{v}_{2,SYG}\left(\hat{Z}_1\right) = \sum\sum_{k<\ell \in s_2^A} \frac{\pi_{1k}\,\pi_{1\ell} - \pi_{1k\ell}}{\pi_{1k\ell}\pi_{2k\ell|s^A}} \left(\frac{z_{2,k}}{\pi_{1k}} - \frac{z_{2,\ell}}{\pi_{1\ell}} \right)^2$, et

$$\mathrm{v}_{\mathrm{SYG}}\left(\hat{Z}_2 \mid s^A\right) = \sum\sum_{k<\ell \in s_2^A} \frac{\pi_{2k|s^A}\pi_{2\ell|s^A} - \pi_{2k\ell|s^A}}{\pi_{1k\ell}\pi_{2k\ell|s^A}} \left(\frac{z_{2,k}}{\pi_{1k}\pi_{2k|s^A}} - \frac{z_{2,\ell}}{\pi_{1\ell}\pi_{2\ell|s^A}} \right)^2 .$$

Dans la phase 1, les probabilités d'inclusion du premier et du second ordre sont respectivement $\pi_{1k} = m / M$ et $\pi_{1k\ell} = \left\{m\left(m-1\right)\right\} / \left\{M\left(M-1\right)\right\}$. Dans la phase 2, les probabilités d'inclusion du premier ordre sont $\pi_{2k|s^A} = n_{2h}^B / n_h^B$ si $k \in s_{2hi}^A$ et $i \in U_h^B$; les probabilités d'inclusion du second ordre dépendent de la strate et de la grappe auxquelles appartiennent les éléments *k* et ℓ. Les différents cas sont les suivants :

i. Même strate *h* et grappe $i \in U_h^B$: $\pi_{2k\ell|s^A} = n_{2h}^B / n_h^B$ si $k, \ell \in s_{2hi}^A$.

ii. Même strate h, mais différentes grappes : $\pi_{2k\ell|s^A} = \left[n_{2h}^B\left(n_{2h}^B-1\right)\right]/\left[n_h^B\left(n_h^B-1\right)\right]$ si i and $i' \in U_h^B$, $i \neq i'$, $k \in s_{2hi}^A$ et $\ell \in s_{2hi'}^A$.

iii. Différentes strates $h \neq h'$, et différentes grappes pour toutes les strates : $\pi_{2k\ell|s^A} = \left(n_{2h}^B n_{2h'}^B\right)/\left(n_h^B n_{h'}^B\right)$ si $i \in U_h^B$, $i' \in U_{h'}^B$, $k \in s_{2hi}^A$ et $\ell \in s_{2h'i'}^A$.

La première partie de $\mathrm{v}_{\mathrm{SYG}}\left(\hat{Z}_2\right)$ de la variance est donnée par $\mathrm{v}_{2,\mathrm{SYG}}\left(\hat{Z}_1\right) = F \sum\sum_{k<\ell\in s_2^A} \frac{1}{\pi_{2k\ell|s^A}}\left(z_{2,k}-z_{2,\ell}\right)^2$, où $F = \frac{M^2\left(1-f^A\right)}{m^2\left(m-1\right)}$ et $f^A = m/M$. Compte tenu des probabilités d'inclusion données plus haut, l'expression de la variance estimée de $\mathrm{v}_{2,SYG}\left(\hat{Z}_1\right)$ peut être simplifiée comme suit :

$$\mathrm{v}_{2,\mathrm{SYG}}\left(\hat{Z}_1\right) = F\left\{\begin{array}{l} \sum_{h=1}^{H}\frac{n_h^B}{n_{2h}^B}\frac{\left(n_h^B-n_{2h}^B\right)}{\left(n_{2h}^B-1\right)}\sum_{i=1}^{n_{2h}^B}\sum_{k\in s_{2hi}^A} m_{hi}\left(z_{2,k}-\bar{z}_{2hi}\right)^2 \\ \qquad + \sum_{h=1}^{H}\frac{n_h^B}{n_{2h}^B}\frac{\left(n_h^B-n_{2h}^B\right)}{\left(n_{2h}^B-1\right)n_{2h}^B} m_{2h}\sum_{i=1}^{n_{2h}^B}\sum_{k\in s_{2hi}^A}\left(z_{2,k}-\bar{z}_{2h}\right)^2 \\ \qquad + \tilde{m}\left[\sum_{h=1}^{H}\frac{n_h^B}{n_{2h}^B}\sum_{i=1}^{n_{2h}^B}\sum_{k\in s_{2hi}^A} z_{2,k}^2 - \frac{1}{\tilde{m}}\left(\sum_{h=1}^{H}\frac{n_h^B}{n_{2h}^B}\sum_{i=1}^{n_{2h}^B}\sum_{k\in s_{2hi}^A} z_{2,k}\right)^2\right] \end{array}\right\} \tag{3}$$

où $\bar{z}_{2hi} = \sum_{k\in s_{2hi}^A} z_{2,k}/m_{2hi}$ et $\bar{z}_{2h} = \sum_{i=1}^{n_{2h}^B}\sum_{k\in s_{2hi}^A} z_{2,k}/\sum_{i=1}^{n_{2h}^B} m_{2hi}$ sont des moyennes pondérées pour les éléments, et $\tilde{m} = \left(\sum_{h=1}^{H}\frac{n_h^B}{n_{2h}^B} m_{2h}\right)$. Il est à noter que cet estimateur est toujours non négatif.

La deuxième partie de l'estimateur de variance $\mathrm{v}_{\mathrm{SYG}}\left(\hat{Z}_2\right)$ est donnée par

$$\mathrm{v}_{\mathrm{SYG}}\left(\hat{Z}_2\middle|s^A\right) = \frac{M^2}{m^2}\sum_{h=1}^{H}\frac{\left(n_h^B\right)^2}{n_{2h}^B}\left(1-\frac{n_{2h}^B}{n_h^B}\right)\frac{1}{n_{2h}^B-1}\sum_{i=1}^{n_{2h}^B}\left(z_{2hi}-\frac{\sum_{i=1}^{n_{2h}^B} z_{2hi}}{n_{2h}^B}\right)^2 \tag{4}$$

où $z_{2hi} = \sum_{k\in s_{2hi}^B} z_{2,k}$.

La version HT de la variance est

$$\mathrm{v}_{\mathrm{HT}}\left(\hat{Z}_2\right) = \mathrm{v}_{2,HT}\left(\hat{Z}_1\right) + \mathrm{v}_{\mathrm{HT}}\left(\hat{Z}_2\middle|s^A\right) \tag{5}$$

La première partie de $\mathrm{v}_{\mathrm{HT}}\left(\hat{Z}_2\right)$ est donnée par

$$\mathrm{v}_{2,HT}\left(\hat{Z}_1\right) = F\left\{\begin{array}{l} \sum_{h=1}^{H}\frac{n_h^B}{n_{2h}^B}\frac{\left(n_h^B-n_{2h}^B\right)}{\left(n_{2h}^B-1\right)}\sum_{i=1}^{n_{2h}^B}\left(z_{2hi}-\frac{1}{n_{2h}^B}\sum_{i=1}^{n_{2h}^B} z_{2hi}\right)^2 \\ \qquad + m\left[\sum_{h=1}^{H}\frac{n_h^B}{n_{2h}^B}\sum_{i=1}^{n_{2h}^B}\sum_{k\in s_{2hi}^A} z_{2,k}^2 - \frac{1}{m}\left(\sum_{h=1}^{H}\frac{n_h^B}{n_{2h}^B}\sum_{i=1}^{n_{2h}^B}\sum_{k\in s_{2hi}^A} z_{2,k}\right)^2\right] \end{array}\right\} \tag{6}$$

Les estimateurs de variance $\mathrm{v}_{2,HT}\left(\hat{Z}_1\right)$ et $\mathrm{v}_{2,SYG}\left(\hat{Z}_1\right)$ sont semblables. Toutefois, $\mathrm{v}_{2,HT}\left(\hat{Z}_1\right)$ peut être négatif, ce qui n'est jamais le cas de $\mathrm{v}_{2,SYG}\left(\hat{Z}_1\right)$. Les estimateurs sont exactement égaux si $\tilde{m} = m$ et si le nombre d'éléments m_{2hi}

dans chaque grappe est le même. La deuxième composante de (5), donnée par $\mathrm{v}_{\mathrm{HT}}\left(\widehat{Z}_2 \middle| s^A\right)$, est la même que pour $\mathrm{v}_{\mathrm{SYG}}\left(\widehat{Z}_2 \middle| s^A\right)$ dans (2).

4. ESTIMATION DE LA COVARIANCE ENTRE p_2 ET p_6

La différence $P_d = P_2 - P_6$ de la population, où $P_2 = Y_2 / M$ et $P_6 = Y_6 / M$, est estimée par $p_d = p_2 - p_6$. On obtient la variance de la différence à l'aide des versions linéarisées de p_2 et p_6 données respectivement par $z^*_{2,k} = (y_{2,k} - P_2)/M$ et $z^*_{6,k} = (y_{6,k} - P_6)/M$. En conditionnant sur s^A et en utilisant les résultats de Tam (1984) sur la covariance d'échantillons chevauchants, on peut montrer que la variance de p_d est donnée par

$$V(p_d) = M^2\left(1 - \frac{m}{M}\right)\frac{S_d^2}{m} + E\left[\frac{M^2}{m^2}\sum_{h=1}^{H} n_h^B\left(\frac{n_{6h}^B}{n_{2h}^B}\widetilde{S}_{2h}^2 + \frac{n_{2h}^B}{n_{6h}^B}\widetilde{S}_{6h}^2 - 2\widetilde{S}_{26h}\right)\right] \tag{7}$$

où $S_d^2 = \dfrac{1}{M-1}\sum_{k=1}^{M}(d_k^* - \overline{D}^*)^2$, $\overline{D}^* = \dfrac{1}{M}\sum_{k=1}^{M} d_k^*$, $d_k^* = z^*_{2,k} - z^*_{6,k}$, et

$$\widetilde{S}_{26h} = \frac{1}{n_h^B - 1}\sum_{i \in s_h^B}\left(z^*_{2hi} - \sum_{i \in s_h^B} z^*_{2hi} / n_h^B\right)\left(z^*_{6hi} - \sum_{i \in s_h^B} z^*_{6hi} / n_h^B\right) \tag{8}$$

avec $z^*_{2hi} = \sum_{i \in s_{2hi}^B} z^*_{2k}$ et $z^*_{6hi} = \sum_{i \in s_{6hi}^B} z^*_{6k}$.

L'estimation de S_d^2 et celle de $\widetilde{S}_{26h}$ exigent que y_{2k} et y_{6k} soient connues pour la même unité, ce qui n'est pas possible, car chaque entreprise remplit une seule version du questionnaire pour chacun de ses employés sélectionnés. On peut utiliser une valeur imputée de y_{2k} pour la comparer à la réponse y_{6k}. Cette méthode a été proposée par Dumais et Lavallée (1990) et, plus récemment, par Van Brakel et Binder (2001). Dans notre contexte, l'ASHE étant une enquête par panel, on génère la valeur en supposant que le schéma de réponse est le même que celui de l'année précédente. Cette valeur étant représentée par y_{2k}^{IMP}, où $k \in s_{6hi}^B$, on peut générer les variables nécessaires pour estimer y_{2k} et y_{6k}. On peut donc maintenant estimer $V(p_d)$ en utilisant le SYG donné par (2) ou la méthode HT donnée dans la section 3.

Les estimateurs de S_d^2 et $\widetilde{S}_{26h}$ exigent les versions linéarisées de p_2 et p_6 définies respectivement par $z_{2,k} = (y_{2,k} - p_2)/\hat{M}_2$ et $z_{6,k} = (y_{6,k} - p_6)/\hat{M}_6$. La différence linéarisée correspondante est $d_k = z_{2,k} - z_{6,k}$. Si l'on utilise la version HT, S_d^2 est estimée par

$$\begin{aligned}\hat{S}_d^2 = \frac{1}{(m-1)}\Bigg\{&\sum_{h=1}^{H}\frac{n_h^B}{n_{6h}^B}\sum_{i=1}^{n_{6h}^B}\sum_{k \in s_{6hi}^B}^{i} d_{hik}^2 - \frac{1}{m}\left(\sum_{h=1}^{H}\frac{n_h^B}{n_{6h}^B}\sum_{i=1}^{n_{6h}^B} d_{hi}\right)^2 \\ &+ \frac{1}{m}\sum_{h=1}^{H}\frac{n_h^B(n_h^B - n_{6h}^B)}{n_{6h}^B(n_{6h}^B - 1)}\left(\sum_{i=1}^{n_{6h}^B} d_{hi}^2 - \frac{\left(\sum_{i=1}^{n_{6h}^B} d_{hi}\right)^2}{n_{6h}^B}\right)\Bigg\}\end{aligned} \tag{9}$$

où $d_{2hi} = \sum_{k \in s_{2hi}^B} d_{2,k}$. De même, l'estimateur de $\tilde{S}_{26h}$ est donné par

$$\tilde{\tilde{S}}_{26h} = \frac{1}{n_{6h}^B - 1} \sum_{i=1}^{n_{6h}^B} \left(z_{2,hi}^{IMP} - \frac{\sum_{i=1}^{n_{6h}^B} z_{2hi}^{IMP}}{n_{6h}^B} \right) \left(z_{6,hi} - \frac{\sum_{i=1}^{n_{6h}^B} z_{6hi}}{n_{6h}^B} \right) \tag{10}$$

où $z_{2hi}^{IMP} = \sum_{k \in s_{2hi}^B} z_{2,k}^{IMP}$ et $z_{6hi} = \sum_{k \in s_{2hi}^B} z_{6,k}$.

5. CONCLUSION

Les calculs de la variance étaient compliqués par la nature intégrée du plan d'enquête et des contraintes connexes. En utilisant le test *t* donné par l'équation (1), nous avons constaté que les taux de réponse étaient légèrement supérieurs pour le questionnaire de six pages que pour le questionnaire de deux pages. En 2005, nous avons ramené le questionnaire de six pages à quatre pages, car cette dernière version était moins coûteuse à traiter. Toutes les questions du questionnaire de six pages ont été transférées au questionnaire de quatre pages. Les résultats obtenus récemment à l'aide du nouveau questionnaire de quatre pages montrent clairement qu'il s'agit d'une amélioration sur le plan de la qualité des données.

RÉFÉRENCES

Dumais, J. et Lavallée, P. (1990), « Une approche pour l'échantillonnage et l'estimation d'enquêtes trimestrielles : l'Enquête sur le camionnage pour compte d'autrui », *Actes du colloque sur les méthodes et domaines d'application de la statistique,* Bureau de la statistique du Québec, Québec, 1990, p. 51 à 58.

Hidiroglou, M.A. et Rao, J.N.K. (2003), « Estimation de la variance dans un échantillonnage à deux degrés », *Actes du Symposium 2003, Défis reliés à la réalisation d'enquêtes pour la prochaine décennie*, Statistique Canada, n° 11-522-XIF au catalogue, Ottawa.

Pont, M. (2003), "Redesigning the UK's Annual Structural Earning Survey", article présenté au *Federal Committee on Statistical Methods*, Washington, D.C., novembre 2003.

Särndal, C.-E., Swensson, B. et Wretman, J. (1992), *Model Assisted Survey Sampling*, Springer-Verlag, New York.

Tam, S.M. (1984), "On Covariances From Overlapping Samples", *The American Statistician*, Vol. 38, No. 4, pp. 288-289.

Van den Brakel, J. et Binder, D.A. (2001), "Variance Estimation for Experiments Embedded in Complex Sampling Schemes", *Proceedings of the Survey Research Methods Section*, American Statistical Association, pp. 805-810.