Techniques d'enquête, décembre 1979
Vol. 5, nº 2
Statistique Canada, nº 12-001-X au catalogue

Échantillonnage avec probabilités inégales et sans remise - une méthode réjective

G.H. Choudhry et M.P. Singh[1]

Résumé

On propose en remplacement de la sélection directe de l'échantillon une autre solution qui, tout en maintenant l'efficacité au même niveau, simplifie les processus de sélection et d'estimation des variances dans un grand nombre de cas. Si n* représente la plus grande taille possible de l'échantillon prélevé selon une méthode qui donne à chaque unité une probabilité d'inclusion proportionnelle à la taille (πPT) à partir d'une population donnée de taille N, la méthode proposée suppose alors la sélection des unités m (= N - n*) en utilisant le schéma πPT et en retirant ces unités de la population de manière à ce que le reste soit un échantillon πPT d'unités n*; l'échantillon définitif des unités n est ensuite prélevé comme sous-échantillon à partir de l'ensemble restant. Cette méthode de sélection de l'échantillon πPT peut être considérée comme l'équivalent de l'EAS dans lequel il est bien connu que la partie « non échantillonnée » de la population et tout sous-échantillon de cette partie constituent également l'EAS de l'ensemble de la population, si l'on applique la procédure EAS. La méthode est très pratique dans les cas où m est inférieur à la taille réelle n de l'échantillon. De plus, elle présente un autre avantage pour les enquêtes permanentes, par exemple l'Enquête sur la population active du Canada (EPA) où il faut augmenter (ou diminuer) le nombre des unités primaires d'échantillonnage (UPE) après la sélection initiale de l'échantillon. La méthode est également intéressante dans le cas du renouvellement de l'échantillon. Le document présente les avantages et inconvénients du plan proposé. L'efficacité de la méthode y est aussi évaluée de façon empirique.

Mots-clés : Échantillonnage avec probabilités proportionnelles à la taille; renouvellement de l'échantillon; efficacité; méthode réjective.

1 Introduction

La sélection d'unités primaires d'échantillonnage (UPE dans un plan d'échantillonnage à plusieurs degrés) avec des probabilités inégales a trouvé de nombreuses applications dans les enquêtes à grande échelle. Cependant, dans de nombreux cas, l'échantillonnage avec remise ou avec une UPE par strate est utilisé en raison de sa simplicité. Les schémas d'échantillonnage sans remise plus efficaces (p. ex. Fellegi [1963], Hartley et Rao [1962], etc.) deviennent assez complexes, qu'il s'agisse de la procédure de sélection ou de l'estimation de la variance, même pour des échantillons de taille modérément grande. Dans le présent article, nous proposons une autre approche (une méthode axée sur le

1. G.H. Choudhry et M.P. Singh, Division des méthodes de recensement d'enquêtes-ménages, Statistique Canada.

rejet) de sélection de l'échantillon qui conserve l'efficacité de la méthode de sélection directe en s'appuyant sur n'importe quel plan d'échantillonnage dans le but de rejeter m unités. Cette méthode simplifie la sélection de l'échantillon et généralement l'estimation de la variance dans une grande diversité de situations. En outre, l'approche suggérée présente plusieurs autres avantages opérationnels dans le contexte des enquêtes continues à grande échelle, notamment :

a) les modifications de l'échantillon, qu'il s'agisse d'augmenter ou de diminuer le nombre d'UPE en fonction des besoins du moment, peuvent être réalisées aussi facilement que dans le cas de l'échantillonnage aléatoire simple (EAS);

b) les bases de sondage à grande échelle sont souvent utilisées comme sources principales de sélection d'échantillons pour des enquêtes ponctuelles de temps à autre (voir Drew, Choudhry et Gray [1978]).

Dans de tels cas, en adoptant cette approche, on peut sélectionner des échantillons à probabilités inégales pour les enquêtes ponctuelles avec la même facilité que l'EAS et sans entrer en conflit avec l'enquête continue principale, une fois que les échantillons ont été sélectionnés pour l'enquête principale.

c) Une caractéristique souvent requise pour les enquêtes continues est la rotation des UPE après un certain temps. Là encore, cette approche permet d'y parvenir facilement, quelle que soit la complexité du schéma de sélection utilisé (pour rejeter m unités).

Cette méthode axée sur le rejet utilisée pour sélectionner des échantillons πPS est très pratique; elle est recommandée surtout dans les situations où la valeur du paramètre m est inférieure à la taille de l'échantillon n, puisque le calcul de π_{ij}, la probabilité que les deux unités i et j se trouvent dans l'échantillon, est en grande partie simplifié. Il convient de souligner que si m est supérieur à la taille de l'échantillon n, il faut utiliser la méthode πPS directe pour sélectionner l'échantillon, à moins qu'il n'y ait des éléments primordiaux à considérer, comme ceux mentionnés précédemment, dans le contexte des enquêtes continues.

Les sections 2 et 3 décrivent le mécanisme de sélection proprement dit et les calculs des probabilités d'inclusion et des probabilités d'inclusion conjointe. Dans la section 4, on démontre la pertinence de cette approche pour la sélection des UPE dans les régions non autoreprésentatives (NAR) de l'Enquête sur la population active (EPA) au Canada. Les résultats de l'étude empirique sont présentés à la dernière section dans laquelle on compare les gains d'efficacité de l'estimateur de Horvitz-Thompson [1952].

2 Procédure de sélection

La population finie donnée U est constituée de N unités, $\{u_1, u_2, \ldots, u_N\}$ et une « mesure de taille » connue x_i est associée à l'unité u_i; i = 1, 2, …, N. Il est nécessaire d'extraire un échantillon de n unités distinctes de la population de telle sorte que la probabilité que l'unité u_i se trouve dans l'échantillon est proportionnelle à sa taille x_i (πPS) pour chaque i = 1, 2, …, N.

Définissons les « tailles normalisées » p_i; i = 1, 2, …, N telles que $\sum_{i=1}^{N} p_i = 1$, c'est-à-dire :

$$p_i = \frac{x_i}{\sum_{i=1}^{N} x_i} \quad ; \quad i=1, 2, \ldots N. \tag{2.1}$$

Un échantillon de n unités distinctes sera sélectionné dans la population de sorte que la probabilité π_i; (i = 1, 2, …, N) que la i^e unité soit dans l'échantillon est np_i. Puisque π_i; i = 1, 2, …, N sont les probabilités et donc nécessairement inférieures ou égales à un, par conséquent, la plus grande valeur possible de n (disons n*) est donnée par :

$$n^* = \left[\frac{1}{p_{(N)}}\right], \tag{2.2}$$

où $p_{(N)} = \text{Max}\ (p_1, p_2, \ldots, p_N)$ et [ ] est la fonction qui prend une valeur entière, c'est-à-dire que la fonction renvoie au nombre entier le plus grand inférieur ou égal à l'argument.

La première étape de la méthode proposée consiste à sélectionner m unités (= N – n*) dans la population en utilisant n'importe quel schéma πPS et à rejeter ces unités de sorte que le reste soit un échantillon πPS de taille n* de la population donnée. Un échantillon aléatoire simple sans remise de n unités retenues sur n* est ensuite sélectionné. Afin de montrer que cet échantillon définitif est un échantillon πPS de taille n de la population entière, nous définissons :

$$p_i^* = \frac{1-n^* p_i}{N-n^*}\ ;\quad i=1,2,\ldots,N. \tag{2.3}$$

On peut facilement vérifier que $p_i^*, (i = 1, 2, \ldots, N)$ sont les probabilités puisque d'après (2.3) $p_i^* > 0$ pour tout i et aussi d'après (2.3) $\sum_{i=1}^{N} p_i^* = 1$, donc, $0 < p_i^* < 1$, $i = 1, 2, \ldots, N$.

Puisque l'échantillon de taille m est sélectionné au moyen d'un schéma πPS donné, la probabilité que l'unité i soit dans l'échantillon est mp_i^*; c'est-à-dire :

$$Pr(i \epsilon R) = mp_i^*,\ i=1,\ 2,\ \ldots,\ N, \tag{2.4}$$

où $R = \{R_1, R_2, \ldots, R_m\}$ est l'ensemble de m unités sélectionnées avec des tailles p_i^*, i = 1, 2, …, N en utilisant un système πPS. Supposons que S* est l'ensemble de n* unités qui ne sont pas dans le R, c'est-à-dire S* = U – R. Maintenant, l'échantillon définitif est un échantillon aléatoire simple de n unités sur n* unités dans s*. Nous désignons par S l'ensemble de n unités sélectionnées lors de la deuxième étape de l'échantillonnage, c'est-à-dire un EAS de n unités sur n* unités dans S*. On peut démontrer que S est un échantillon πPS de taille n de la population U de taille N comme suit :

$$\Pr(i \in S^*) = \Pr(i \notin R)$$

$$= 1 - mp_i^* ; \qquad i=1, 2, \ldots, N. \tag{2.5}$$

En substituant $m = N - n^*$ et p_i^* de (2.2) dans l'équation (2.5), on obtient :

$$\Pr(i \in S^*) = n^* p_i ; \qquad i=1, 2, \ldots, N. \tag{2.6}$$

Ainsi, S* est un échantillon πPS de taille n* (la plus grande taille d'échantillon autorisée) de U. Puisque S est l'EAS de n sur n* dans S*, nous avons :

$$\Pr(i \in S) = \Pr(i \in S^*) \times \Pr(i \in S \mid i \in S^*)$$

$$= (n^* p_i) \times \left(\frac{n}{n^*}\right)$$

$$= np_i \quad ; \qquad i=1, 2, \ldots, N. \tag{2.7}$$

En désignant par π_i, la probabilité que l'unité i soit dans l'ensemble S, on écrit :

$$\Pi_i = np_i \quad ; \qquad i=1, 2, \ldots, N, \tag{2.8}$$

c'est-à-dire que S est un échantillon πPS de taille n de la population U de taille N.

Lorsque $m << n$, l'échantillon de m unités peut être sélectionné en utilisant, par exemple, la méthode de Fellegi [1963] ou la méthode d'échantillonnage systématique avec classement aléatoire et PPT (probabilité proportionnelle à la taille) imputable à Hartley et Rao [1962] en fonction de la valeur du paramètre m et de la taille de la population. N. Sinha [1973] a également proposé un schéma d'échantillonnage avec rejet qui permet d'obtenir des probabilités d'inclusion préétablies des deux premiers ordres. La pertinence d'autres méthodes peut également être étudiée pour diverses situations.

3 Calcul de π_{ij}

La formule pour π_{ij}, la probabilité que les unités i et j se trouvent toutes deux dans l'échantillon désigné par S dans la section précédente, sera obtenue. En désignant par π^{*}_{ij}, la probabilité que les unités i et j se trouvent toutes deux dans S*, on obtient immédiatement :

$$\Pi_{ij} = \frac{n(n-1)}{n^{*}(n^{*}-1)} \times \Pi^{*}_{ij}; \quad \begin{matrix} i=1, 2, \ldots, N-1, \\ j=i+1, i+2, \ldots N. \end{matrix} \tag{3.1}$$

Afin de trouver π^{*}_{ij}, nous définissons les quatre événements suivants, exhaustifs et mutuellement exclusifs, pour les unités i et j par rapport aux ensembles R et S*.

Événement E1 : les deux unités i et j dans R.

Événement E2 : l'unité i dans R et l'unité j dans S*.

Événement E3 : l'unité i dans S* et l'unité j dans R.

Événement E4 : les deux unités i et j dans S*.

Alors nous obtenons :

$$\begin{aligned} \Pi^{*}_{ij} &= Pr(E4) \\ &= 1-\{Pr(E1) + Pr(E2) + Pr(E3)\} . \end{aligned} \tag{3.2}$$

mais

$$\begin{aligned} Pr(E1) + Pr(E2) &= Pr(i \epsilon R) \\ &= mp^{*}_{i} . \end{aligned} \tag{3.3}$$

De même

$$\begin{aligned} Pr(E1) + Pr(E3) &= Pr(j \epsilon R) \\ &= mp^{*}_{j} . \end{aligned} \tag{3.4}$$

L'addition de (3.3) et (3.4) donne :

$$Pr(E1) + Pr(E1) + Pr(E2) + Pr(E3) = m(p_i^* + p_j^*). \qquad (3.5)$$

En substituant dans (3.2) pour Pr(El) + Pr(E2) + Pr(E3) à partir de (3.5) nous obtenons :

$$\begin{aligned} \Pi_{ij}^* &= 1 - m(p_i^* + p_i^*) + Pr(E1) \\ &= 1 - m(p_i^* + p_i^*) + \delta_{ij} \end{aligned} \qquad (3.6)$$

où

$$\begin{aligned} \delta_{ij} &= Pr(E1) \\ &= Pr(\text{Both } i, j \in R). \end{aligned}$$

En substituant π_{ij}^* de (3.6) ci-dessus dans (3.1), on obtient π_{ij}, la probabilité que i et j se trouvent dans l'ensemble S, c'est-à-dire :

$$\Pi_{ij} = \frac{n(n-1)}{n^*(n^*-1)} \times [1 - m(p_i^* + p_j^*) + \delta_{ij}];$$

$$i = 1, 2, \ldots, N-1,$$
$$j = i+1, i+2, \ldots, N, \qquad (3.7)$$

où $m = N - n^*$ et p_i^*, $(i = 1, 2, \ldots, N)$ est définie dans (2.3).

Dans l'échantillonnage traditionnel (ou direct), plus n augmente, plus le calcul de π_{ij} devient compliqué, mais le cas échéant, la complexité réside uniquement dans le calcul de δ_{ij}, qui dépend de la valeur du paramètre m. Étant donné que m est un paramètre de population et ne dépend pas de la taille de l'échantillon, la complexité de π_{ij} n'augmentera pas avec la taille de l'échantillon. On peut mentionner que, pour le cas particulier où la valeur du paramètre m est égale à 1, $\delta_{ij} = 0$, donc :

$$\begin{aligned} \Pi_{ij}^* &= 1 - (p_i^* + p_i^*) \\ &= (N-1)(p_i + p_j) - 1 \end{aligned} \qquad (3.8)$$

et

$$\Pi_{ij} = \frac{n(n-1)}{(N-1)(N-2)} [(N-1)(p_i+p_j)-1]. \qquad (3.9)$$

Puisque dans ce cas, une seule unité (m = 1) sera rejetée avec PPT, aucun plan d'échantillonnage spécial n'est requis avec cette méthode axée sur le rejet.

4 Application à l'Enquête sur la population active au Canada

L'Enquête sur la population active (EPA) au Canada suit un plan d'échantillonnage stratifié à plusieurs degrés [voir Statistique Canada, n° 71-526 au catalogue]. Dans les régions non autoreprésentatives (NAR), constituées de régions rurales et de petits centres urbains, un échantillon πPS d'UPE est sélectionné dans chaque strate, où la « mesure de la taille » est la population totale de l'UPE du recensement précédent. Dans la conception antérieure, avant le remaniement de 1971, la méthode de Fellegi [1963] était utilisée pour sélectionner les UPE, alors que deux UPE étaient sélectionnées dans chaque strate. Cette méthode, bien qu'assez efficace, devient très compliquée lorsque le nombre d'UPE à sélectionner est important, généralement plus de trois. Lors du remaniement de l'EPA à la suite du Recensement de 1971, l'un des critères de choix de la procédure de sélection au stade de la conception était que la procédure devait être suffisamment souple pour permettre l'expansion de l'échantillon pour ce qui est du nombre d'UPE ainsi que la rotation des UPE. La méthode d'échantillonnage systématique avec classement aléatoire et PPT (probabilité proportionnelle à la taille) [Hartley et Rao, 1962] a été adoptée pour sélectionner les UPE, car elle répond aux exigences nécessaires [Gray, 1973]. Cependant, le calcul de π_{ij} devient compliqué pour les grandes valeurs de n et N. Le schéma d'échantillonnage proposé, bien que tout aussi efficace, présente l'avantage supplémentaire de la simplicité pour l'expansion de l'échantillon et le calcul de π_{ij}.

À la suite du remaniement de l'EPA, l'échantillon a été élargi pour ce qui est du nombre d'UPE dans les régions NAR et le nombre actuel d'UPE sélectionnées par strate varie d'une province à l'autre. Le nombre actuel d'UPE sélectionnées dans chaque strate au sein d'une province et le nombre de cas avec m < n par province sont indiqués dans le tableau A1 (annexe A). Dans le contexte de la simplification des calculs des probabilités d'inclusion conjointe pour l'estimation de la variance, on constaterait que sur 127 strates NAR au Canada, 108 donneraient lieu à des calculs plus simples, 16 à une difficulté égale et 3 seulement à une plus grande difficulté. Puisque le schéma proposé a l'avantage de permettre d'augmenter l'échantillon en sélectionnant simplement des UPE supplémentaires avec l'EAS, la procédure, entre autres, peut être considérée pour la sélection des UPE dans les régions NAR lors du remaniement de l'EPA à la suite du Recensement de 1981. De plus, après l'augmentation de l'échantillon, les nouveaux π_{ij} seront reconstruits à partir des précédents en les multipliant simplement par le facteur constant approprié.

5 Étude empirique

Nous avons choisi quatre populations pour évaluer de façon empirique le schéma d'échantillonnage avec rejet pour des échantillons de taille 2, 3 et 4. La description de ces populations est donnée dans le tableau 1, où x est la valeur connue de la mesure de la taille, et y est la valeur de la caractéristique d'intérêt qui est inconnue, mais mesurable. La valeur du paramètre m est également indiquée dans le tableau 1 pour chacune des populations.

Tableau 1
Description des populations pour l'étude empirique

Pop.	N°	Source	N	m	CV(x)[1]	CV(y)	ρ_{xy}
1		Fellegi [1963]	6	2	0,25	0,64	0,93
2		Gray [1971 b]	10	2	0,08	0,07	0,93
3		Cochran [1963, p. 204].	10	3	0,17	0,19	0,97
4		Cochran [1963, p. 225].	10	2	0,14	0,15	0,65

[1] $CV(x) = \dfrac{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2 / N}}{\bar{x}}$

CV(y) est défini de la même manière.

Dans la présente étude, le schéma d'échantillonnage avec rejet est réalisé en rejetant m unités avec la méthode avec PPT de Fellegi [1963] et également en rejetant m unités avec la méthode d'échantillonnage systématique avec classement aléatoire et PPT de Hartley et Rao [1962]. On considère l'estimateur de Horvitz et Thompson [1952] $\hat{Y}_{HT} = 1/n \sum_{i\varepsilon S} y_i/p_i$, où $\sum_{i\varepsilon S}$ désigne la somme des n unités de l'échantillon, pour estimer le total inconnu de la population $y = \sum_{i=1}^{N} y_i$. La variance de $\hat{Y}_{HT}$ telle qu'elle est donnée par Yates et Grundy [1953], c'est-à-dire

$$V(\hat{Y}_{HT}) = \frac{1}{n^2} \sum_{i<j} \sum (\pi_i \pi_j - \pi_{ij}) \left(\frac{y_i}{p_i} - \frac{y_j}{p_j}\right)^2 \qquad (5.1)$$

a été calculée pour le schéma avec rejet lorsque m unités sont rejetées avec la méthode avec PPT de Fellegi et également lorsque m unités sont rejetées avec la méthode d'échantillonnage systématique avec classement aléatoire et PPT; ces variances ont ensuite été comparées à celles de l'échantillonnage direct avec la méthode avec PPT de Fellegi et la méthode d'échantillonnage systématique avec classement aléatoire et PPT, respectivement. Afin de calculer les probabilités conjointes π_{ij} pour la méthode avec PPT de Fellegi, les « probabilités de travail » de la méthode ont été calculées par une procédure itérative et les π_{ij} ont été construites en additionnant les probabilités de tous les échantillons qui contiennent à la fois les unités i et j. Pour des tailles d'échantillon supérieures à 4, la procédure devient plus compliquée et suppose un grand nombre de calculs fastidieux. Pour la méthode d'échantillonnage systématique avec classement aléatoire et PPT, les π_{ij} ont été calculées à l'aide d'un sous-programme FORTRAN de Hidiroglou et Gray [1979]. L'algorithme utilisé par les auteurs susmentionnés est une modification de la formule de Connor [1966] et est imputable à Gray [1971a]. Les variances pour la méthode avec rejet s'appuyant sur la méthode de Fellegi pour rejeter m unités et pour la méthode de Fellegi pour sélectionner l'échantillon des échantillons de taille 2, 3 et 4 sont données dans le tableau B1 (annexe B). Une comparaison semblable est faite en remplaçant la méthode de Fellegi par la méthode systématique avec classement aléatoire et PPT tant pour le rejet de m unités dans la

méthode avec rejet que pour la sélection de l'échantillon; ces résultats sont présentés dans le tableau B2 (annexe B).

Selon les deux tableaux de l'annexe B, le niveau d'efficacité de la méthode avec rejet est le même que celui de la méthode avec PPT utilisée pour rejeter les m unités, c'est-à-dire la méthode avec PPT de Fellegi et la méthode systématique avec classement aléatoire et PPT. De plus, une comparaison entre les tableaux B1 et B2 montre que les variances de la méthode avec PPT de Fellegi et de la méthode systématique avec classement aléatoire et PPT sont les mêmes, et sont donc également efficaces. Puisque l'efficacité de la méthode avec rejet est la même que celle de la méthode avec PPT utilisée dans la méthode avec rejet pour rejeter m unités, il sera avantageux d'utiliser la méthode avec rejet pour les populations où la valeur du paramètre m est inférieure ou égale à la taille réelle de l'échantillon n.

Remerciements

Les auteurs souhaitent remercier l'examinateur pour ses commentaires utiles.

Bibliographie

[1] Cochran, W.G., (1963), *Sampling Techniques, Second Edition*, New York: John Wiley & Sons, Inc.

[2] Connor, W.S. (1966), "An Exact Formula for the Probability that Two Specified Sampling Units Will Occur in a Sample Drawn with Unequal Probabilities and Without Replacement", *Journal of the American Statistical Association*, 61, 384-390.

[3] Drew, J.D., Choudhry, G.H. et Gray, G.B. (1978), "Some Methods For Updating Sample Survey Frames and Their Effects on Estimation", *Techniques d'enquête*, Statistique Canada, 4, 225-263.

[4] Fellegi, I.P. (1963), "Sampling With Varying Probabilities Without Replacement: Rotating and Non-Rotating Samples", *Journal of the American Statistical Association*, 58, 183-201.

[5] Gray, G.B. (1971), "Joint Probabilities of Selection of Units in Systematic Samples", *Proceedings of the American Statistical*.

[6] Gray, G.B. (1971), "Variance Components and Variance Functions", Comptes rendus de "Statistique Canada" 71, 119-126.

[7] Gray, G.B. (1973), "On Increasing the Sample Size (number of PSU's)", Note technique interne, Census and Household Survey Methods Division, Statistique Canada.

[8] Hartley, H.O. et Rao, J.N.K. (1962), "Sampling with Unequal Probabilities and Without Replacement", *Annals of Mathematical Statistics*, 33, 350-374.

[9] Hidiroglou, M.A. et Gray, G.B. (1979), "Construction of Joint Probability of Seleciton for Systematic PPS Sampling", *Journal of the Royal Statistical Society*.

[10] Horvitz, D.G. et Thompson, D.J. (1952), "A Generalization of Sampling without Replacement from a Finite Universe", *Journal of the American Statistical Association*, 47, 663-685.

[11] Sinha, B.K. (1973), "On Sampling Schemes to Realize Pre-assigned Sets of Inclusion Probabilities of First Two Orders", *Calcutta Statistical Association Bulletin*, 22, 89-110.

[12] Statistique Canada (1977), "Methodology of the Canadian Labour Force Survey (1976)", n° 71-526 au catalogue.

[13] Yates, F. et Grundy, P.M. (1953), "Selection without Replacement from Within Strata with Probability Proportional to Size", *Journal of the Royal Statistical Society, Series B*, 15, 235-261.

Annexe A

Tableau A1

Nombre de strates ayant un échantillon avec rejet d'unités par rapport au nombre réel d'unités sélectionnées dans les strates non autoreprésentatives de l'Enquête sur la population active au Canada

		nombre de de strates avec m <, = ou > n			nombre de strates avec m égal à n			
Province	**n**	**m < n**	**m = n**	**m > n**	**1**	**2**	**3**	**4**
Terre-Neuve-et-Labrador	4	11	1	0	6	5	2	1
Île-du-Prince-Édouard	6	3	0	0	0	3	0	0
Nouvelle-Écosse	3	9	5	1	3	6	5	1
Nouveau-Brunswick	3	10	4	0	2	8	4	0
Québec	3	18	1	2	3	15	1	2
Ontario	3	15	5	0	3	12	5	0
Manitoba	6	7	0	0	4	3	0	0
Saskatchewan	6	11	0	0	2	5	4	0
Alberta	4	14	0	0	9	5	0	0
Colombie-Britannique	4	10	0	0	4	5	1	0
Canada		108 difficulté moindre[1]	16 difficulté égale	3 difficulté supérieure	34	67	22	4

[1] La difficulté moindre, égale ou supérieure fait référence aux calculs des probabilités d'inclusion conjointe.

Annexe B

Tableau B1

Variance pour le schéma avec rejet s'appuyant sur la méthode de Fellegi pour rejeter m unités (schéma 1) et variance pour la méthode de Fellegi pour sélectionner l'échantillon (schéma 2)

		Taille de l'échantillon		
Pop. N°	**Schéma d'échantillonnage**	**2**	**3**	**4**
1	Schéma 1	8,0161	3,6782	1,5092
	Schéma 2	8,1672	3,8258	1,5269
	Efficacité (1 par rapport à 2)	101,88%	104,01%	101,17%
2	Schéma 1	3,4922	2,0475	1,3251
	Schéma 2	3,4948	2,0509	1,3287
	Efficacité (1 par rapport à 2)	100,07%	100,17%	100,27%
3	Schéma 1	276,04	161,67	104,48
	Schéma 2	276,15	161,81	104,63
	Efficacité (1 par rapport à 2)	100,04%	100,09%	100,14%
4	Schéma 1	6 375,5	3 756,6	2 447,1
	Schéma 2	6 373,2	3 753,7	2 444,1
	Efficacité (1 par rapport à 2)	99,96%	99,92%	99,88%

Tableau B2

Variance pour le schéma avec rejet s'appuyant sur la méthode systématique avec classement aléatoire et la méthode avec probabilité proportionnelle à la taille (PPT) pour rejeter m unités (schéma 3) et variance pour la méthode systématique avec classement aléatoire et PPT pour sélectionner l'échantillon (schéma 4)

		Taille de l'échantillon		
Pop. N°	**Schéma d'échantillonnage**	**2**	**3**	**4**
1	Schéma 3	8,0261	3,6915	1,5242
	Schéma 4	8,5073	4,3927	1,5242
	Efficacité (3 par rapport à 4)	106,00%	119,00%	100,00%
2	Schéma 3	3,4922	2,0475	1,3251
	Schéma 4	3,5114	2,0423	1,3309
	Efficacité (3 par rapport à 4)	100,55%	99,75%	100,44%
3	Schéma 3	276,41	162,16	105,03
	Schéma 4	276,20	160,23	103,65
	Efficacité (3 par rapport à 4)	99,92%	98,81%	98,69%
4	Schéma 3	6 376,6	3 758,0	2 448,7
	Schéma 4	6 373,7	3 750,7	2 446,2
	Efficacité (3 par rapport à 4)	99,95%	99,81%	99,90%